VISIT…



Microscopy Techniques for Materials Science

A R Clarke and C N Eberhardt

CRC

WP

Microscopy techniques for materials science

A R Clarke and C N Eberhardt



CRC Press
Boca Raton Boston New York Washington, DC

WOODHEAD PUBLISHING LIMITED
Cambridge England

Published by Woodhead Publishing Limited
Abington Hall, Abington
Cambridge CB1 6AH
England
www.woodhead-publishing.com

Published in North America by CRC Press LLC
2000 Corporate Blvd, NW
Boca Raton FL 33431
USA

First published 2002, Woodhead Publishing Limited and CRC Press LLC
© 2002, Woodhead Publishing Limited
The authors have asserted their moral rights.

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. Reasonable efforts have been made to publish reliable data and information, but the authors and the publishers cannot assume responsibility for the validity of all materials. Neither the authors nor the publishers, nor anyone else associated with this publication, shall be liable for any loss, damage or liability directly or indirectly caused or alleged to be caused by this book.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming and recording, or by any information-storage or retrieval system, without permission in writing from the publishers.

The consent of Woodhead Publishing Limited and CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from Woodhead Publishing Limited or CRC Press LLC for such copying.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

Woodhead Publishing ISBN 1 85573 587 3
CRC Press ISBN 0-8493-1552-2
CRC Press order number: WP1552

Cover design by The ColourStudio
Project managed by Macfarlane Production Services, Markyate, Hertfordshire (macfarl@aol.com)
Typeset by MHL Typesetting Limited, Coventry, Warwickshire
Printed by TJ International Limited, Padstow, Cornwall, England

To our wives and children

Sue, Hannah, Gemma, Emily and Rosalie Clarke

and

Sue Eberhardt

Contents

Preface ix

A*cknowledgements* xiii

List of Figures xvii

List of Tables xxix

Part I. Basic Principles

1. Interaction of EM Radiation with Materials 3
   - 1.1 Introduction 3
   - 1.2 Characteristics of EM Radiation 9
   - 1.3 Propagation of Light Waves 24
   - 1.4 Elements of Microscope Designs 47
   - 1.5 Photonics 73
   - 1.6 References 84
   - 1.7 Bibliography 85
2. Digital Imaging and Processing 86
   - 2.1 Introduction 86
   - 2.2 Digital Data 88
   - 2.3 The History of Digital Computing 92
   - 2.4 Charge Coupled Devices (CCDs) 110
   - 2.5 Digitisation and ADCs 116
   - 2.6 Digital Images 125
   - 2.7 Storage and Retrieval of Images 131
   - 2.8 Image Enhancement 137

This page has been reformatted by Knovel to provide easier navigation.

2.9 Errors and Stereology 156
2.10 References 159

Part II. 2D Optical Reflection and Confocal Laser Scanning Microscopy

3. 2D Optical Reflection Microscopy 163
3.1 Introduction 163
3.2 A Large Area Automated, High Spatial Resolution Microscope (LAAM) 179
3.3 Case Study: Fibre Orientations within Injection-Moulded Composites 199
3.4 Case Study: the Measurement of Fibre Lengths in Image Fields 212
3.5 Latest Developments 218
3.6 References 224
4. 3D Confocal Laser Scanning Microscopy 228
4.1 Principles of Confocal Laser Scanning Microscopy 228
4.2 Modern CLSMs 235
4.3 Optical Sectioning Capability of the CLSM 240
4.4 Calibration Issues 249
4.5 Imaging Modes 254
4.6 Case Study: Thin Film Particulate Analysis 267
4.7 Case Study: Fibre Waviness 278
4.8 Future Possibilities for CLSM 298
4.9 References 298

Part III. Other Microscopical Techniques

5. Complementary Optical and EM Imaging Techniques 305
5.1 Introduction 305
5.2 Raman Microscopy 307
5.3 Scanning Probe Microscopy/Near-Field Microscopy 314

This page has been reformatted by Knovel to provide easier navigation.

5.4 Miscellaneous Optical and EM Techniques 322
5.5 X-Ray Microscopy/Tomography/ Microtomography 331
5.6 Case Study: X-Ray Microtomography of Fibrous Structures 337
5.7 References 353
6. Other Microscopy Techniques 358
6.1 Electron Microscopy (SEM/TEM) 361
6.2 Nuclear Magnetic Resonance (NMR) 371
6.3 Ultrasonics and Scanning Acoustic Microscopy 376
6.4 Case Study: Ultrasonic Mapping of 3D Stiffness Constants of Composite Materials 380
6.5 Epilogue 404
6.6 References 406

Index **410**

This page has been reformatted by Knovel to provide easier navigation.

Preface

At last the book is finished – and I have now been asked to put my mind to the Preface! It occurs to me that writing a Preface is a unique art form. Admittedly, after limited research into Preface-writing, I propose, like innumerable authors before me, to start with the usual whinge – yes, to paraphrase Mrs Beeton from the Preface of her famous cookbook, if we had known '… what courageous efforts were needed to be made', I am quite sure that we would never have started this enterprise. However, it is clear that one of the sensible reasons for co-authorship is that it has, at least, halved the agony for each of us.

In a sense, our book could be considered a kind of 'cook-book' – a cook-book for any reader who is interested in image processing and microscopy as a means to a materials science research end. Over the past two years, the original form of the book, as previously discussed with Patricia Morrison in 1999, has changed dramatically, but we hope, for the better. We have tried to minimise the mathematics that underpins these topics, and have concentrated on the practical issues (and pitfalls) one comes across when acquiring and analysing image data using various microscopic and tomographic measurement techniques. We have also made a real effort to expunge all spelling and grammatical *faux pas* and believe that, scientifically, the equations are correct and the conclusions presented (such as they are) are sound. In view of the speed of evolution in microscopic measurement techniques, this book can only hope to be a snapshot of the current situation and we suggest that readers follow the latest research articles in journals like the Royal Microscopical Society's *Journal of Microscopy* to keep abreast of future developments. Also, you might like to keep checking our Department of Physics and Astronomy website here at the University of Leeds to see our latest 3D reconstructions.

Recently, a colleague of mine in the Molecular Physics and Instrumentation Group said that he did not have a clue what I had been doing (research-wise) for the past ten years – well, this book is for him. In effect, the various case studies, in the latter two thirds of this book, catalogue my research team's attempts to try to understand how to make the best measurements of fibre orientations and their spatial distributions in both composites and textiles materials. They provide

testimony to just how far we have progressed from the initial 2D image analyser (which was attached to a standard optical reflection microscope), via the confocal laser scanning microscopes (Biorad MRC500 and Noran Olympus) and finally onto the X-ray microtomography and ultrasonic 'time of flight' systems (whose full potential we are still exploring).

One of the most difficult decisions we had to make was where to start the book. We eventually decided, without wishing to insult the intelligence of the reader, that a good place to start would be to review some background topics. Therefore, the first part of the book sets the scene for what later follows and is a rapid canter through historical developments in both optics and electronics, which have naturally led to the current computer-assisted, microscopy techniques that exist today.

We hope you like our sprinkling of apt quotations. Most books that deal with science, computers or computing are often incomprehensible and tedious. It is a pity that the surrealism of Lewis Carroll's works cannot be sustained throughout a 'scientific' book like this one. I have been looking for literary allusions to microscopy and, as might be expected, Carroll has something appropriate to say (aimed at children) in 'The Professor's Lecture' from one of his lesser well-known works, *Sylvie and Bruno Concluded*:

> … he beckoned the Gardener to come up on the platform, and with his help began putting together what looked like an enormous dog-kennel, with short tubes projecting out of it on both sides.
>
> 'But we've seen elephants before,' the Emperor grumbled.
>
> 'Yes, but not through a ***Megaloscope***!' the Professor eagerly replied. 'You know you can't see a Flea, properly, without a ***magnifying-glass*** – what we call a ***Microscope***. Well, just in the same way, you can't see an Elephant properly without a ***minimifying-glass***. There's one in each of these little tubes. And this is a Megaloscope! …''

Lewis Carroll then goes on to describe a hilarious scene with the megaloscope, and a second passage from the same Professor's Lecture also touches upon the central theme of our book.

> … 'Our Second Experiment', the Professor announced, as Bruno returned to his place, still thoughtfully rubbing his elbows, 'is the production of that seldom-seen-but-greatly-to-be-admired phenomenon, ***Black Light***! You have seen White Light, Red Light, Green Light, and so on: but never, till this wonderful day, have any eyes but mine seen ***Black Light***! This box', he said, carefully lifting it upon the table, and covering it with a heap of blankets, 'is quite full of it … would anyone like to get under the blankets and see it?'
>
> Dead silence followed this appeal: but at last Bruno said 'I'll get under …'
>
> … 'What did you see in the box?' Sylvie eagerly enquired.
>
> 'I saw nuffin!' Bruno sadly replied. 'It were too dark!'
>
> 'He has described the appearance of the thing exactly!' the Professor exclaimed with enthusiasm.

> ***Black Light***, and ***Nothing***, look so extremely alike, at first sight, that I don't wonder he failed to distinguish between them! We will now proceed to the Third Experiment. …'

Perhaps, with good reason, he was also taking the rise out of the university lecturing profession here! Having quoted already from Lewis Carroll, it seems appropriate that I conclude this hastily conceived piece with some good advice that Carroll has to offer in his Preface to the *Sylvie and Bruno Concluded* story:

> … Let me here express my sincere gratitude to the many Reviewers who have noticed, whether favourably or unfavourably, the previous Volume. Their unfavourable remarks were, most probably, well-deserved; the favourable ones less probably so. Both kinds have no doubt served to make the book known, and have helped the reading Public to form their opinions of it. Let me also here assure them that it is not from any want of respect for their criticisms, that I have carefully forborne from reading ***any*** of them. I am strongly of the opinion that an author had better ***not*** read any reviews of his books: the unfavourable ones are almost certain to make him cross, and the favourable ones conceited; and ***neither*** of these results is desirable …

So, dear reader, if (for some inexplicable reason) you do not like our book or find any glaring errors, do not bother to tell us, just use the book as an expensive door-stop. If you do like the book, please do not bother to let us know, but for sure, tell all your friends and colleagues about it, because both Dr Eberhardt and I have expanding families to support and any extra royalties would be most welcome.

Ashley Clarke, September 2002

Acknowledgements

I have indeed been fortunate to be the supervisor of some very talented people, whose special software skills, physical insight and attitude towards their research work has resulted in our development of the measurement techniques described by the case studies in this book. I owe a great debt to my Ph.D students over the past 12 years: Dr Nic Davidson for the development of the 2D image analyser system; Dr Geoff Archenhold for his work on the Biorad MRC500 confocal system and the initial work into pattern matching; Dr Mike Enderby for automating the first version of the ultrasonic testrig and especially my co-author, Dr Colin Eberhardt for extending the confocal work with the Noran Odyssey and, latterly, the X-ray microtomography research. Also, I would like to thank my current Ph.D students: Andrew Schwarz, who has helped tremendously with the necessary corrections to some of the figures through his knowledge of the Adobe Illustrator software package (and his digital photography of some of the test equipment within the department) and Mat Harper for his input to the final case study on 3D elastic stiffness constants and the ultrasonic testrig.

For our recent foray into the X-ray microtomography research, my special thanks go to Nishanth Gopinathan and Dr Jia, the Institute for Particle Science and Engineering, School of Process, Environmental and Materials Engineering, for their assistance with, and access to, the University of Leeds, Skyscan 1072 X-ray system. Also, Professor Ryszard Pyrz at the Institute of Mechanical Engineering, Aalborg University for many years of fruitful collaboration, exchange of postgraduate students and access to their Skyscan machine.

With a small team like ours, the research contribution from short-term, undergraduate project students can be significant and we have been fortunate with visiting French project students over the years: Georges Bervin, Dan Gutknecht, Noe Poffa and Homig Lamon. There have also been numerous Leeds 3rd year project students who were persuaded to contribute to our confocal and ultrasonic work, amongst them were Sze Wei Ku, Andrew Johnston, James Watt and Kathryn Morris. We have also been assisted by French postgraduate students from Dr Michel Vincent's group at the École des Mines; Thomas Giroud and Sylvain Fluoret.

Special thanks go to our colleagues in the Polymer IRC and Polymer Group at the University of Leeds: Dr Alan Duckett, Dr Peter Hine and Professor Ian Ward who have vastly more experience than I with the measurement and modelling of fibre reinforced polymer composites. Their interaction with us over the years has sparked off many useful developments and I am indebted to Dr Duckett for steering me onto the problem of fibre orientations over 13 years ago during a memorable coffee break in 1989! I am also indebted to our colleagues in the Molecular Physics and Instrumentation Group within the Department of Physics and Astronomy: Professor David Batchelder, Dr Alastair Smith, Dr Simon Webster, Dr Kurt Baldwin and Ph.D student Kevin Critchley for their help on Raman microscopy which has made a valuable contribution in this book. My special thanks also go to Dr Mike Ries for the NMR photograph in Chapter 6.

I would like to now take the opportunity to acknowledge the skills of our Mechanical and Electronic Workshops (led by Mr Jack Coley and Mr Mansukh Patel respectively) within the Department of Physics and Astronomy, especially the skills of Trevor Haines, Stewart Weston, Andrew Price and Paul Ogden who have all contributed directly to our ultrasonic testrig design over the years.

On the practical side of microscopy, my special thanks go to Dr Vyvyan Howard (University of Liverpool) for introducing me to the mysteries and potential of stereology, Dr Torsten Mattfeldt for collaborating with us on a test for isotropy within fibre-reinforced composites and Dr Alan Entwistle (Ludwig Institute and the Royal Microscopical Society) for his many useful discussions and invaluable contribution to team funds through his annual microscope workshop demonstrations!

Many organisations have supplied us with samples over the years, some of which figure in this book, but my particular thanks go to: Cranfield University (Professor Phil Irving), IKP University of Stuttgart (Dr Gunther Fischer), École des Mines, Sophia-Antipolis (Dr Michel Vincent), University of Bristol (Professor Mike Wisnom), ICI Technology (Dr Bill Meredith and Dr Simon Allen) who have all provided us with a broad range of samples (ranging from bread, wood adhesive, foam and thin films), DuPont (Dr Paul Mills) and the Leeds School of Textile Industries (Dr Stephen Russell).

Like most small research teams, the financial assistance needed to keep together a team and its collective expertise has always been a struggle. My especial thanks go to Professor Paul Curtis (DERA Farnborough) for large and small research contracts granted to us over the years; also our unseen sponsors in Brussels, who were persuaded to fund our work under the Brite-Euram initiative; the UK Science Research Council for the numerous MoD/EPSRC joint grants that have been awarded to us over the past 10 years, and the assistance of liaison officer, Dr Matthew Hiley on the latest grants. Without the timely assistance of Dr Andrew Dickson (Biorad) and his free loan of the Biorad MRC500 confocal system for a number of years, we could never have performed our initial CLSM

studies on composites. Then, without Noran UK's half-price offer of an Olympus CLSM system and Dr Garry Burdett's (Health and Safety Executive, Sheffield) significant contribution towards the Olympus, we would never have been able to explore the automation of confocal systems. My co-author (CNE) is also grateful for being awarded ICI Technology's Case Quota Ph.D Studentship on thin films.

Our grateful thanks go to Dr Bill Meredith for providing the phase contrast, DIC and crossed polar images shown in Figures 1.62 and 1.63; Professor Tony Wilson (University of Oxford) for permission to show his tilted microcircuit in Figure 4.5; Professor Gwynne Morgan (Leeds) for his photograph of Sir William Bragg (Figure 5.1); Dr Kurt Baldwin and Dr Simon Webster for the Raman figures (Figures 5.6, 5.7 and 5.14) and LEO Electron Microscopy Ltd for the picture of the first commercial electron microscope in Figure 6.3. Other figures in the book have been recreated, from many referenced sources, by Colin Eberhardt using Adobe Illustrator software. Nearly all of the figures in most of the case studies have been generated from our own image data using the Leeds 2D system, the Noran Odyssey CLSM, the Skyscan 1072 system or the Leeds ultrasonics testrig.

Finally, our thanks to Gwen Jones, Patricia Morrison and Stuart Macfarlane for keeping us on schedule.

Ashley Clarke, September 2002

Figures

1.1	(a) (b) and (c) Indication of nanoscale through microscale to mesoscale features	6
1.2	Techniques versus ultimate axial and lateral resolution	7
1.3	(a) Overview of image processing functions (b) overview of stereological ideas	8
1.4	Finite Element modelling mesh	9
1.5	A typical 3D visualisation technique – glass fibres in polymer composite	10
1.6	Abstraction of wavelike properties – a mathematical model	11
1.7	(a) and (b) Illustration of shear and tensile waves in crystal structures	12
1.8	Electric and magnetic field variations in space	13
1.9	Electromagnetic spectrum	14
1.10	Ray trace of incident, reflected and refracted waves at interfaces	16
1.11	Fibre optics – trapping of light waves	17
1.12	Refractive index versus wavelength – showing absorption bands in UV	18
1.13	Group velocity – modulation envelope	19
1.14	(a) Linear polarisation (b) elliptical (c) circular polarisation (left/right)	20
1.15	Lissajous figures on oscilloscope screen	21
1.16	(a) Birefringence (b) linear polariser + quarter wave plate	23
1.17	(a) Diffuse reflection (b) specular reflection	25
1.18	Blooming of a lens surface-destructive interference between reflected beams to minimise the overall reflectance from a lens surface	26
1.19	(a) and (b) TE/TM modes of incidence on interface	27
1.20	Brewster's angle condition	28
1.21	Variation of external/internal reflection with incident angle	29
1.22	Images with spherical mirror (a) and (b) different object locations	30

1.23	Basis of interference effect between two coherent beams	31
1.24	Young's slit and the intensity pattern shown schematically on a screen where $\theta = y/L$ and $d =$ distance of point P off-axis and $L =$ distance between the screen and slits (slit separation $d = 1.5\,\text{mm}$, and slit width $a = 6\mu\text{m}$)	32
1.25	Michelson interferometry arrangement	33
1.26	Transmission of optical filters (a) BG3 type glass filter response, (b) band pass and high pass (c) beamsplitter and barrier filter	34
1.27	Neutral density filter responses	35
1.28	(a) Light spread outside geometric shadow (b) Huygens wavelets construction for diffraction	36
1.29	Intensity distribution through the bright and dark fringes gives this diffraction pattern due to single slit (slit width 9 microns)	36
1.30	Diffraction pattern due to circular aperture (Airy pattern)	37
1.31	Overview of absorption processes	38
1.32	Black-body radiation curves – UV catastrophe, and Wien's Law	40
1.33	Photoelectric effect	40
1.34	Emission lines of excited hydrogen	41
1.35	Electron waves in allowed orbits	42
1.36	Illustration of different processes by electron jumps (a), (b), (c), (d), (e), (f) and (g)	43
1.37	Schematic diagram of the operation of a pm tube with dynode chains	44
1.38	Edison lamp (from Edison's patent 'Electric-Lamp', Jan 27 1880, #223,898)	46
1.39	(a) Schematic diagram of laser (b) laser process	47
1.40	(a) Cross-section of the eye and (b) image on the retina	48
1.41	Ray-tracing of image for a convex lens and lens formula	49
1.42	Ray-tracing of image for a concave lens	50
1.43	Variation in magnification/position of object	51
1.44	(a) and (b) Distinguishing two point sources – Airy patterns overlapping	52
1.45	Influence of diffracted orders through an objective lens	53
1.46	Telescope design	54
1.47	Compound Microscope design	55
1.48	Oil immersion objective lens with coverslip	57
1.49	(a) Longitudinal chromatic aberration (b) doublet lens	59
1.50	Spherical aberration effect	60
1.51	(a) and (b) Oil immersion plan-apochromat and achromat multiple lenses	61
1.52	Achromat and apochromat colour correction	62
1.53	Critical illumination overview	62
1.54	Koehler illumination overview	63

1.55 Darkfield condenser design 65
1.56 Reflecting optics microscope design 66
1.57 (a) Basic schematic and (b) episcopic fluorescence microscope 68
1.58 Transforming phase variations into intensity variations 69
1.59 Phase contrast microscope 70
1.60 Wave front shearing interference microscope 71
1.61 (a) Wollaston prism (b) Nomarski Interference microscope 72
1.62 (a) and (b) Darkfield image and phase contrast 74
1.63 Cross-polar images of red ink 75
1.64 Kerr cell 77
1.65 Pockels cell 77
1.66 Diffracting aperture and spatial filter 78
1.67 High pass/low-pass filters 79
1.68 Holographic-based systems 81
1.69 Spatially-switched systems and vector matrix multiplier 81
1.70 Lenslet array and spatial light modulator (SLM) 82
1.71 Smallest confocal microscope design 84

2.1 (a) AND and OR functions using switches (b) AND and OR functions using semiconductor diodes 91
2.2 Transistor in logic gates – DTL and TTL 93
2.3 (a) Flip-flops and (b) different types of RAM: SRAM, DRAM 94
2.4 Basis of the Read-Only Memory 96
2.5 Dual Port/VRAM – Toshiba TC528267 97
2.6 Overview of a 6802 microprocessor chip 98
2.7 Overview of the VELA, a typical 6802 microprocessor-based system 101
2.8 Instruction codes routing data within 6802 CPU 102
2.9 (a) Transputer T800 overview (b) some parallel processing architectures: pipeline, mesh array and hypercube 104
2.10 The A100 DSP chip 106
2.11 (a) RC lowpass filter and (b) low-pass filter frequency response 107
2.12 A100 implementation of the low-pass filter 108
2.13 Evolution of computer power, adapted from [11] 110
2.14 How a CCD works: 3-step process 111
2.15 Spectral responses of CCD devices, (a)–(e) 114
2.16 Operating temperatures of various CCD devices 115
2.17 General purpose ADC and interaction with host computer 117
2.18 Fast flash-ADC (3 bit version) 118
2.19 Slow scan creation of an image 120
2.20 Staircase of digital values (a) no offset (b) offset by q/2 121
2.21 Typical line signal in video format 121
2.22 Nyquist sampling theorem illustration 122

2.23	C80 digital signal processor overview	123
2.24	Overview of the Genesis framegrabber board	125
2.25	(a) A continuous function which has been (b) digitised	126
2.26	The storage of a digital image	127
2.27	Spatial aliasing – a diffraction grating at a number of different resolutions	128
2.28	The image processing route – overview of processes	129
2.29	The structure of a Windows Bitmap image data file	133
2.30	Image compression	136
2.31	The effect of JPEG on a 3×3 Laplace transform: (a) and (b) uncompressed (c) and (d) JPEG compressed	138
2.32	A number of images of the same region demonstrating how frame averaging improves the quality of the image	140
2.33	A reflection mode image of a mirror clearly demonstrates the non-uniform image gain over the field of view. Intensity contours are plotted for pixel intensity steps of 10 units.	141
2.34	Histogram equalisation transformation	142
2.35	Histogram manipulation of images	143
2.36	Edge enhancing convolution filter example (a) original image (b) filtered	145
2.37	(a) Baboon image (b) FFT (c) high-pass and (d) low-pass	147
2.38	FFT enhances periodic features. Sample: 5 micron latex spheres dispersed using a solvent which has evaporated – used as a mask for deposition of metallic dots	148
2.39	(a) An image which has been subjected to regular spatial noise, (b) its Fourier transform and (c) the inverse FFT after the peaks corresponding to the spatial noise have been removed	149
2.40	(a) Image plane (b) the corresponding curves in parameter space	150
2.41	Applications of the Hough transform to machine vision	151
2.42	Circular Hough transform applied to bubble image	153
2.43	Segmentation of features by the application of an intensity threshold	154
2.44	An illustration of the recursive pixel connectivity algorithm showing the iterative ‘flood fill’ of a segmented region. The images run sequentially from (a) to (e).	155
2.45	Size bias due to edge effects	158
2.46	An unbiased counting frame	159
3.1	A cut-away of an Olympus BH2 transmission mode design	164
3.2	Mesoscale image of gear wheel tooth	165
3.3	Microscale size image of a composite showing elliptical fibres and fibre fragments	167
3.4	Photograph of the Struers Rotopol-11 polishing machine	168

3.5	Sectioned samples potted in epoxy resin block	169
3.6	Photo of sputtering rig at Leeds for oxidation of surfaces	169
3.7	Photo of an Olympus BH2 microscope with XY stage	170
3.8	A schematic of a generalised fibre and co-ordinate axes	171
3.9	A typical pixellated near-circular image to illustrate intrinsic digital noise	172
3.10	Effect of pixel aspect ratio on error $\delta\theta$ in angle, θ	173
3.11	Finding the best pixel aspect ratio from 2D data sets	174
3.12	Typical POM dataset showing the $\theta = 15^{o}$ peak even when the fibres are isotropically distributed	174
3.13	Solid angle weighting effect	175
3.14	Height weighting effect	175
3.15	Illustration of θ, ϕ transformation for large angle sections	176
3.16	Kidney bean-shaped cross-section carbon fibres	177
3.17	Typical stealth film image using a flatbed scanner	180
3.18	An overview of the hardware for the Leeds 2D image analyser	181
3.19	Network of transputers in original image analyser design	182
3.20	Operation of stepper motors	184
3.21	Cyclic variations in the linear movements in X and Y	185
3.22	Raster scanning with overlap to show how larger areas are covered	186
3.23	A set of images to illustrate the creation of a collage	187
3.24	Using second moments technique to determine ellipticities	188
3.25	Least squares technique to derive ellipticity of fibre fragments	189
3.26	Typical screen dump showing (θ, ϕ) plots, the current frame and the previous frame	190
3.27	Yurgatis 'included angles' characterisation of structure	193
3.28	$K(r)$ – the second order intensity function	194
3.29	Effect of flood-filling on hollow glass fibres, (a) (b) and (c)	195
3.30	Surface image of triangular fibres (cut at 90 degrees)	196
3.31	Schematic of a plane at (α, β) intersecting a triangular fibre	196
3.32	Relationship between θ_{ab} and α, β	197
3.33	Triangular cross-section fibres which have been sectioned at 45° to main fibre orientation direction	198
3.34	Histogram of the fibre orientations after sectioning at 45 and 90 degrees to main triangular fibre orientation direction	198
3.35	Reflected light microscope image of a fibre reinforced composite sample potted in epoxy	200
3.36	The ellipticity of a fibre cross-section depends on its orientation with respect to the section plane	201
3.37	The relationship between fibre orientation and the major and minor axes (and the ambiguity on determining the angle ϕ)	201
3.38	Classification of objects shown in a scatter plot	204

3.39	Orientation dependence on sampling showing stereological bias	205
3.40	Two touching fibre cross-sections with their edges highlighted and the 8-pixel tracing route shown	206
3.41	Illustration of how curvature is measured	207
3.42	An image with a group of touching fibres which have been split	208
3.43	3D probability distribution of fibre orientations	209
3.44	Three examples of orientation distributions (i) unidirectional (ii) isotropic (iii) unidirectional (at 45º to coordinate axes)	210
3.45	(a) Conventional (b) 8-stroke injection moulded samples	211
3.46	Stress profile for a long and short fibre (30%vf E-Glass in Nylon 66 resin) calculated using the shear lag model (H.L. Cox, 1952)	213
3.47	Transmission mode image of fibres that remain after the surrounding matrix has been pyrolised. Note the presence of both crossing fibres and connected fibres within the field of view	214
3.48	(a) The co-ordinate system used to characterise the fibre's position and orientation and (b) clusters in (θ, ρ) space	215
3.49	Crossing fibres and the derivation of perimeter angle	217
3.50	Curvature signatures for crossing fibres	217
3.51	The length of a curved fibre is measured by summing the lengths of a number of short straight line segments	218
3.52	Fibre length measurement for a 5 mm section of a unidirectional composite	219
3.53	Fibre length distribution within the skin and core regions of an injection moulded composite	219
3.54	Illustration of Paluch scheme and the fibre straightening effect	220
3.55	For the cluster technique, the offset between section planes	221
3.56	Cluster of fibre images to give a control point	222
3.57	Location of clusters along the large area diagonal line	223
3.58	Schematic of McGrath & Wille optical sectioning technique	224
4.1	Portrait of Marvin Minsky and the earliest confocal designs	229
4.2	Four optical sections of a pollen grain each separated by a distance of $\Delta z = 10\,\mu$m	230
4.3	A schematic of a confocal laser scanning microscope (CLSM), from Åslund *et al.*	232
4.4	PSFs: conventional microscope and confocal microscope	234
4.5	Images of tilted microcircuits, from Wilson	235
4.6	Axial response for a confocal microscope	236
4.7	Schematic of folded optics for the Biorad MRC500	237
4.8	(a) Optical schematic of the Noran 'Odyssey' CLSM (b) the complete CLSM system at Leeds with Odyssey and Prior XYZ stage	239

4.9	The axial resolution is characterised by $\Delta z_{1/2}$, which is plotted in terms of the numerical aperture of three popular types of objective lens	241
4.10	(a) The axial response of the Noran Odyssey CLSM with various aperture sizes and (b) the axial response of the fluorescent light path with various slit sizes	242
4.11	(a) An oil immersion objective lens imaging without aberration and (b) with spherical aberration due to refractive index mismatch	243
4.12	Contour plots of the aberrations in the confocal PSF as an oil immersion lens is focussed through a sample with the same RI as water (from Hell *et al.*)	244
4.13	A volume dataset of particulates in a thin polymer film	245
4.14	Low packing fraction and high packing fraction, fluorescence XZ images	247
4.15	A plot of the pixel fluorescence intensity as a function of depth showing the exponential nature of the attenuation of the fluorescence signal	248
4.16	A reflection mode image of a mirror clearly demonstrates the non-uniform image gain over the field of view. Intensity contours are plotted for each intensity step of 10 units	250
4.17	Inclined mirror produces the axial response function	250
4.18	Line profiles through an image of a mirror at two different gain settings. Dotted lines show the line profile after background shading correction	251
4.19	Magnification variations across the field of view	253
4.20	A diagram showing the interpolation of pixel intensities from the stretch corrected image to the original image	254
4.21	The excitation and emission spectra for fluorescein dye	255
4.22	(a) An image of a polyurethane foam sample with labelled particulates. The scale bar represents a length of 1 mm (b) plot of particulate penetration within the foam	256
4.23	(a) Reflection mode and (b) fluorescence mode image of the surface of a glass fibre reinforced, polymer composite sample. Scale bar represents 20 μm	257
4.24	An XY frame showing the photobleached lines due to previous XZ sections. Scale bar represents 20 μm	258
4.25	A schematic diagram illustrating two methods of projection	259
4.26	Volume rendering of composite	260
4.27	Volume rendering of a polymer foam sample	260
4.28	XY sections at different Z	261
4.29	XZ sections at different Y	262
4.30	A depth map of a silicon chip	263

4.31 The repeatability of the z-drive is indicated by repeatedly acquiring a surface profile of a mirror slide 263
4.32 A schematic representation of an objective lens focusing along a surface with complex topology 264
4.33 A reflection mode surface image of a group of metallised latex spheres. The scale bar represents a length of 20 μm 265
4.34 (a) Surface height profile and (b) 3D rendered image of the surface topology of a thin polymer film 266
4.35 An auto-focus image of a polyurethane foam sample, taken using Nikon PlanApo $\times$10 air objective lens. This image was acquired over a range, $\Delta z = 300\,\mu$m 267
4.36 A schematic illustration of the industrial production of polymer film by extrusion and bi-axial drawing 268
4.37 The disector principle. (a) reference section and (b) look-up section 269
4.38 The unbiased brick counting frame 270
4.39 Intensity artefacts in particulate detection 272
4.40 The intensity histogram and volume (in voxels) at different gain settings 273
4.41 A flow chart demonstrating the combination of objects 274
4.42 A 'composite' volume reconstruction created from six datasets acquired at different PMT gain settings 275
4.43 (a) A cross section through a single particulate (taken from the dataset illustrated in Figure 4.39) (b) the same region, binarised by the application of a single threshold (c) the original cross section smoothed by a 3x3 mean filter and (d) the binarised, smoothed image 275
4.44 Unbiased counting frame 276
4.45 Particulate density distributions 277
4.46 Number weighted frequency distribution 278
4.47 Volume weighted frequency distribution 278
4.48 (a) Unidirectional loading (one fibre misaligned) (b) and (c) kink band formation (d) failure 280
4.49 Compressive failure shown by mesoscale image 281
4.50 Cross-section image showing fibre waviness 282
4.51 CLSM images taken at increasing depths produce movement of fibre centres 283
4.52 XZ sections of composite (a) in fluorescence and (b) in reflection mode 284
4.53 A schematic of the XZ sectioning approach used to reconstruct fibre waviness 285
4.54 A noisy XZ image which has been thresholded and segmented 286
4.55 The calculation of the local image intensity gradient G 287

4.56 Iterative fibre centre location search 288
4.57 A number of plies containing both fibre and matrix are heated and compressed (in the direction of the arrows) to form the final composite. The axes illustrate the orientation at which the sample was mounted under the microscope 288
4.58 An 'extended' XZ section taken using a × 60, NA 1.4, Nikon PlanApo objective lens. The image has dimensions of 280 μm×50 μm in XZ 289
4.59 The entire sampled volume, 440 μm×5000 μm×50 μm in XYZ, containing 472 fibre segments. There has been a considerable compression of the illustrated volume in the Y direction 290
4.60 Fibre length distribution for the volume reconstruction given in Figure 4.59 290
4.61 Fibre misalignment angle frequency, with fibre orientation angles calculated in the plane of the ply (in-plane) and perpendicular to the plane of the ply (out-of-plane) 291
4.62 Fitting polynomials of different order to the raw fibre data 293
4.63 A frequency histogram of the uncertainty in fibre centre location 294
4.64 The percentage error in the mean radius of curvature of the fitted space curve for simulated fibres with a length of 2 mm 295
4.65 The optimum polynomial order for the fitted space curve for a number of fibre segment lengths 295
4.66 The smoothed dataset from Figure 4.59 296
4.67 The radius of curvature distribution for the composite sample and the simulated straight fibre dataset with instrumental noise 296
4.68 Fibre waviness bias introduced when a technique has a limited depth of penetration 297

5.1 Sir William Henry Bragg and his early X-ray crystallography experimental rig 306
5.2 Bragg spectrometer apparatus used for teaching and research 307
5.3 Schematic of CH_2 molecular vibrations which give rise to infrared and Raman lines 308
5.4 Stokes and Anti-Stokes lines 309
5.5 Photograph of the Ramanscope 310
5.6 Schematic of the optical train of the Ramanscope 311
5.7 Typical spectral scans showing pressure effect on Raman lines 312
5.8 Raman effect used to determine the variation of strain within a fibre 314
5.9 A typical force–distance curve experienced by a probe tip as it approaches a surface 315
5.10 A schematic overview of all SPM techniques 316
5.11 Photograph of a typical STM used in Leeds for research 318

5.12 Photograph of the original AFM built by Binnig and Rohrer 319
5.13 Schematic showing the basis of near-field microscopy 321
5.14 UV RS SNOM (Leeds design) schematic 322
5.15 Thermography technique for detecting defects 323
5.16 FT-IR systems – (a) dispersive and (b) interferometric designs 324
5.17 (a) Mechanical arrangement for a balloon-borne lamellar grating interferometer (b) a typical interferogram 326
5.18 Principle behind FLIM 327
5.19 FLIM system overview 328
5.20 PAM system overview 329
5.21 Optical coherence tomography system overview 330
5.22 Typical curves of the mass absorption coefficients for different atomic number target materials and various X-ray energies 331
5.23 X-ray generation method and X-ray spectrum 332
5.24 Zone plate theory 334
5.25 Berglund's apparatus for soft X-ray research 336
5.26 The famous radiograph made by Roentgen on 22 December 1895, which is traditionally known as 'the first X-ray picture' 338
5.27 During X-ray tomography, parallel X-rays pass through the sample resulting in the production of a shadow image. The dotted line indicates the section reconstructed by back projection in Fig. 5.28 339
5.28 The reconstruction of a single sample cross-section by back-projection 340
5.29 Common artefacts present in CT images 342
5.30 Measurement of fibre orientation using Hough transform 344
5.31 Reconstruction of three different materials 346
5.32 (a) Parameters used in the fibre determination routine (b) voxel intensity as function of fibre segment orientation 348
5.33 Steps for joining the most likely fibre segments 349
5.34 Three 3D reconstructions of different fibrous systems 350
5.35 Derivation of fibre orientation tensor coefficients 351
5.36 Frequency distribution of fibre curvature for yarn sample using space curve approach 351
5.37 Frequency distribution of fibre torsions for yarn sample using space curve approach 352

6.1 Basis of use of magnetic and electrostatic lenses for electron beam refraction and focussing 359
6.2 Comparison of basic operation of standard optical microscope, electrostatic version of electron microscope and magnetic coil version of electron microscope 360
6.3 Photograph of first commercial electron microscope 361

6.4	Photograph of modern scanning electron miscoscope	362
6.5	Schematic overview of different products created by electron beam hitting a material	365
6.6	(a) Typical spectra for the XPS/ESCA technique (b) schematic of photo-ionisation process and (c) schematic of oxygen KLL Auger process	367
6.7	Energy spectra of back-scattered electrons	368
6.8	Schematic of the high resolution EELS microscope	369
6.9	(a) Energy sensitivity of the EELS system (b) electron energy-loss spectrum of an oxide superconductor	370
6.10	Visualisation of the Zeeman and NMR effects and magnetic moment concept	372
6.11	Typical NMR system for research at Leeds	373
6.12	NMR spin-echo scheme	374
6.13	Schematic of acoustic waves at liquid/solid interface	377
6.14	The ZnO transducer and associated fused quartz lens of a SAM system	378
6.15	Different types of pulsed ultrasonic systems	380
6.16	3D stresses on a small volume element due to applied forces	382
6.17	Photograph of the Leeds ultrasonic testrig	384
6.18	Schematic of Leeds TOF testrig	385
6.19	Oscilloscope plots (a) transmitted and received pulse trains (b) received pulse train showing time resolution	386
6.20	Angles of incidence, reflection and refraction for ultrasonic waves propagating through a parallel-sided plaque	387
6.21	Variation of ultrasonic wave velocity as function of temperature	388
6.22	Typical line scan at normal incidence for ultrasonic wave	388
6.23	Attenuation of ultrasound signal as a function of thickness of a Perspex block	390
6.24	Typical TOF area scan for Perspex sample	391
6.25	Correlation between TOF and signal size for Perspex at normal incidence	391
6.26	Variation of (a) TOF with angle of incidence and (b) signal size with angle of incidence for Perspex sample	392
6.27	Measurement of all 3D stiffness constants within a parallel-sided plaque	393
6.28	Shear wave TOF variations at a fixed point on a unidirectional, carbon fibre reinforced sample as fibre orientations rotated	395
6.29	(a) TOF map for normal incidence and (b) TOF map for shear waves (circular insert in unidirectional carbon sample)	396
6.30	Perpendicular line scans of TOF across the circular carbon fibre insert at (a) normal incidence and (b) 40º incident angle	397

6.31 Correlated TOF and intensity data for shear waves at incident beam angle of 35º 398
6.32 (a) Injection moulded, glass fibre-reinforced plaque and (b) variation in fibre orientation tensor coefficients across sample thickness 399
6.33 Variation of TOF at different positions on injection-moulded, glass fibre reinforced polypropylene plaques 400
6.34 Correlation plots of TOF against signal size for one of the samples shown in Fig. 6.33 401
6.35 Stiffness constant maps derived from a number of TOF area scans taken at different angles of incidence 402
6.36 Frequency distributions of the stiffness constant values 403

Tables

1.1	Different types of waves and their velocities in materials	13
1.2	Refractive indices of materials (at wavelength of 550 nm)	15
1.3	Sign conventions for thin lens formula	51
1.4	Depth of field values for different objectives	58
1.5	Working distances of different objective lenses	64
2.1	The representation of the numbers 10–15 in hexadecimal	89
2.2	The matrix of alphanumeric characters (ASCII)	90
2.3	OR, AND, NOR, NAND, EX-OR, EX-NOR truth tables, symbols and gates	91
2.4	Bitmap Info Header of a Microsoft Windows™ bitmap	134
3.1	Internal angles for three fibres	197
4.1	Working distances of typical CLSM objectives	246
4.2	Refractive indices of polymer matrices	285

Part I

Basic principles

1

Interaction of EM radiation with materials

1.1 Introduction

It might be thought that the use of optical microscopes for materials science research was self-evident and the last thing that was required was yet another book. However, such is the pace of modern technological developments (in computing and photonics) that we believe there is a niche for a book, which discusses both instrumentation issues and the potential of the latest optical microscopes for materials science research.

The role of the microscopist is essentially to interpret the 2D or, better still, the 3D structure of the specimen being studied. Therefore, surely all that the microscopist needs to do, when faced with a new specimen to analyse, is to:

- select an appropriate light microscopical technique
- physically section the specimen (and possibly polish the surface)
- analyse the features – usually features in 2D image field(s)
- compute suitable parameters which describe effectively these features in 3D.

What could be simpler? The problems, however, begin to show when one considers each of these operations in more detail. Firstly, there are many variations on a theme (even for optical microscopy) and the user must be clear what type of measurement is most likely to answer the research questions that are posed. For example, what is the appropriate scale size of the features that need to be investigated? The answer to this question no doubt depends on the rationale behind the microscopical analysis, and also the type of material will affect the choice of technique. Secondly, which physical section of the sample should be taken to be representative of the whole material? If the material is homogeneous and isotropic, the section taken is irrelevant, but any anisotropy within the sample (for example in fibre-reinforced composites) could cause serious problems for the interpretation of the true structure. Thirdly, a suitable surface finish for the optical technique selected may require more detailed preparation of the surface than simply polishing – etching and sputtering might be required, for example! Finally, it is one thing to interpret the 2D cross-section

features of 3D objects within the sample and another matter to describe adequately the 3D objects without experimental bias creeping into the calculations. Yet another crucial issue to be addressed is 'how much data must be gathered to obtain a statistically significant result?'

When this book was commissioned, the brief was to try and create a 'how to do it' book where the emphasis was on case studies and practicalities, rather than dwelling on the substantial mathematics which underpins all branches of this subject. Hence, to appeal to a wider range of potential readers, the book has been organised into three main sections.

Part I is really for beginners to either microscopy or the computer processing of images. More mature readers may wish to skip over Chapter 1. This chapter sets the scene by discussing some fundamental ideas behind the propagation of electromagnetic radiation and its interaction with materials and also basic issues in optical microscopy. However, there are a number of topics which might merit a cursory read through. Chapter 2 describes briefly the development of electronics which gave rise to microprocessors and, through the concepts of digital signal processing and computing paradigms, to the latest image framegrabbers. This chapter also shows how computer-assisted microscopy generates digital images and discusses various ways in which images can be transformed to pick out the features of interest to the microscopist.

Part II considers two of the most common and important techniques for optical studies of materials. Chapter 3 concentrates on the 2D optical reflection microscope technique, as typified by the system designed at Leeds, and there are a number of case studies on its applications to fibrous material samples, e.g. fibre-reinforced composites and textiles. Chapter 4 deals with the confocal laser scanning microscope (CLSM), and more 3D reconstruction case studies are discussed including thin film research. The case studies developed in Part II use basic ideas from Part I (e.g. selection of a software approach to optimise the analysis followed by the interpretation of the materials data), and illustrate how complete solutions to a materials problem can be formulated.

Part III illustrates other experimental (mainly optical) techniques in materials science. Chapter 5 deals with variations on a theme: alternative optical and EM techniques, which are proving important nowadays. Finally, Chapter 6 covers, briefly, acoustic and a few other techniques being used successfully in materials science research.

The main objective for computer-assisted microscopy is to make the most effective measurements of the 3D structure of materials – to make the measurements speedily and efficiently and to interpret the resulting images correctly. The authors hope that this book will illustrate the types of measurement that are now possible and point the way to future uses of novel optical microscopy techniques.

As promised above, the book begins with an exploration of the basic physics behind the propagation of light in space and through materials. The

schizophrenic world of the physicist is explored where light can be considered as having both wave and particle-like properties (i.e. light being a stream of discrete entities called photons). However, dear reader, if you are impatient and wish to pass over the rest of Part I, the following three subsections provide an overview of some key issues from Chapters 1 and 2.

1.1.1 Process – structure – properties

Defining the scope of one's problems is perhaps the hardest task in a research programme. In materials science, the important issues for the microscopist probably relate to the reliability of a manufacturing process for a component part, which meets the required design strength specifications. This is essentially a quality control function. Conventional parts will have dimensions ranging from many centimetres through to many metres, but nowadays research is gradually leading to the manufacture of nanometre (10^{-9} metre) scale size component parts too. The range of sample dimensions to be investigated therefore now covers nanoscale through the microscale and on to the mesoscale regime, as shown in Fig. 1.1.

The impact of glass fibre and carbon fibre reinforced composites on the development of high strength and low weight materials over the past 50 years has generated a need for more computer-assisted microscopy. Now the use of textiles as the reinforcing component in polymer composites or fibrous structures for medical applications has added an extra challenge to 3D reconstruction of these materials. In the recent past, researchers have tried to improve the processing techniques by measuring the resulting large-scale material properties from variations in the processing parameters. Clearly, the importance of the structural studies is that one can have a better understanding of why a material fails or why one processing route is better than another route, if good quality 2D or 3D measurements are available.

1.1.2 Optical and non-optical techniques

Another role for materials science is to characterise the 3D structure of the material under study. Hence, if the material is heterogeneous (i.e. composed of different spatial structures), the spatial extent of those structures, Δl (in x, y or z), must be probed by some form of electromagnetic radiation (or acoustic pressure waves) whose wavelength, λ, is a fraction of Δl. Figure 1.2 gives an overview of the limits of spatial resolution for different techniques of measurement to be found in Parts II and III.

Since the time of Lord Rayleigh in the nineteenth century until the mid/late-twentieth century, it was assumed that light microscopical techniques would never beat the resolution limit dictated by the wavelength of the light illumination used (see Section 1.4.1). Only in the past twenty years has it been

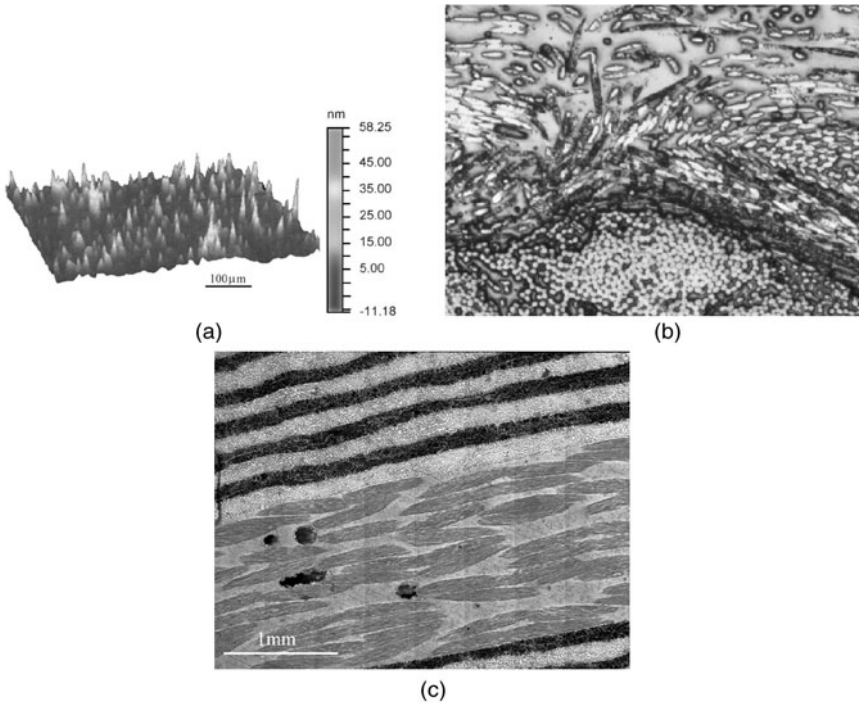


Figure 1.1 (a) An illustration of nanoscale size effects ($10^{-9} - 10^{-6}$ m) is shown in this WYKO profilometer study of unfilled PET film. (b) A typical microscale size image ($10^{-6} - 10^{-3}$ m) is illustrated by a complex, fibre-reinforced composite. (c) A mesoscale view ($10^{-3} - 10^{-1}$ m), created from a collage of high resolution images, is shown for a cross-section through a helicopter blade.

possible to image individual atoms within a crystal structure and to reach nanometre scale resolution with devices like the scanning tunnelling microscope (STM) and scanning near field microscopy (which are discussed in the final chapter).

1.1.3 Computer-aided functionality to microscopes

There are a number of ways by which computers add functionality to conventional microscopes. Firstly, the computer speeds up the acquisition and archiving of raw image data. Secondly, the computer allows for the processing of those images in order to improve the contrast (pick out specific features better) and minimise noise (by filtering and averaging operations). Thirdly, the computer allows the user to interpret the image data and reduce operator errors (for example when counting objects or estimating lengths of objects). The 3D interpretation of the 2D image data is also enhanced by stereological methods

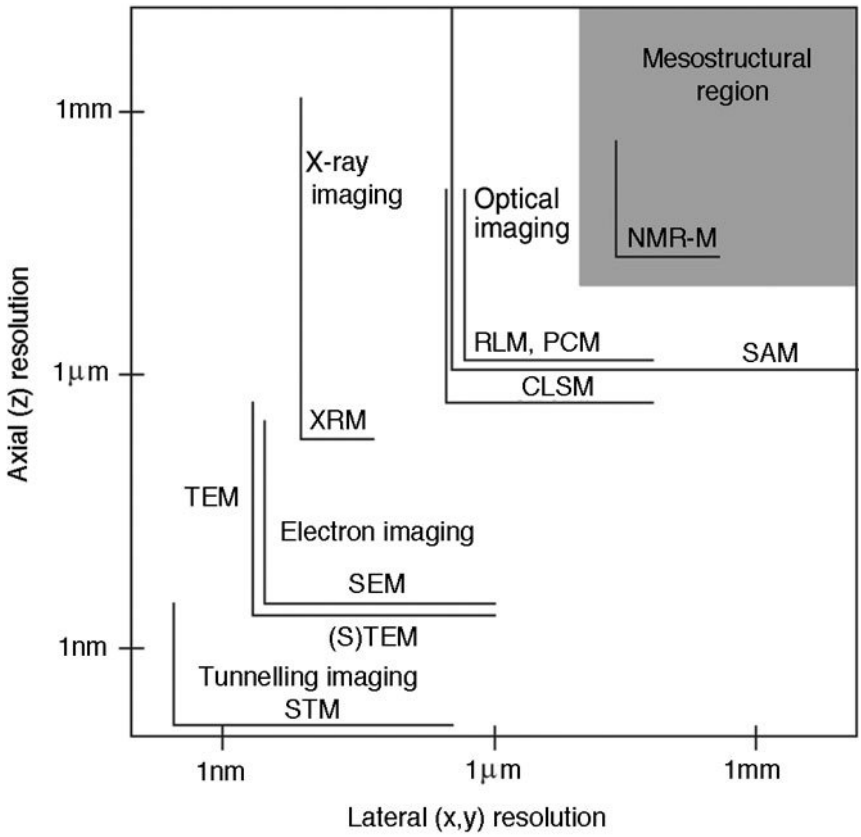


Figure 1.2 The current limits to spatial resolution for the different techniques discussed in this book: CLSM (confocal laser scanning microscopy), SAM (scanning acoustic microscopy), NMR (nuclear magnetic resonance), SEM (scanning electron microscopy), STM (scanning tunnelling microscopy).

requiring repetitive operations to be performed, which is ideal for the computer. An overview of some image processing and stereological methods is shown in Fig. 1.3.

1.1.4 Modelling and 3D visualisation issues

Finally, ever-increasing computer processing power can be applied to the mathematical modelling, which relates either processing conditions to final part structures or which relates the observed part structure to final product properties. One of the most popular modelling methods in use today is finite element analysis (see Fig. 1.4). Mathematically, the structure to be analysed is subdivided into a mesh of finite sized elements of simple shape. Stresses can be applied to the simulated sample and the resulting strains are propagated from

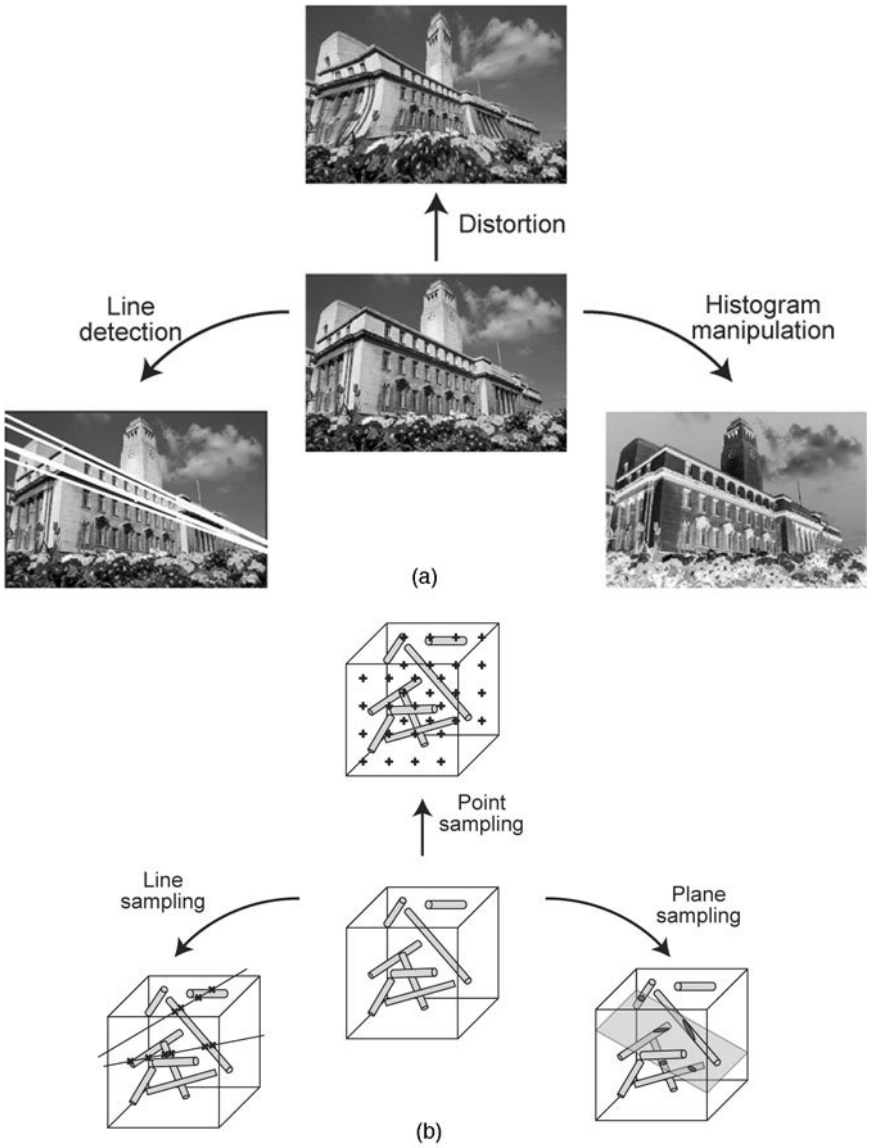


Figure 1.3 (a) After digitisation of the image, many different processes may be used to transform/modify the image in order to detect linear features, alter the contrast to make features more visible or to distort the image for startling visual effects. (b) Stereology seeks to define unbiased methods for estimating feature characteristics using appropriate sampling probes. These probes could be point probes to measure numbers, or line probes or 2D sampling planes, usually with associated special counting rules. Great care must be exercised when interpreting 2D images of heterogeneous structures.

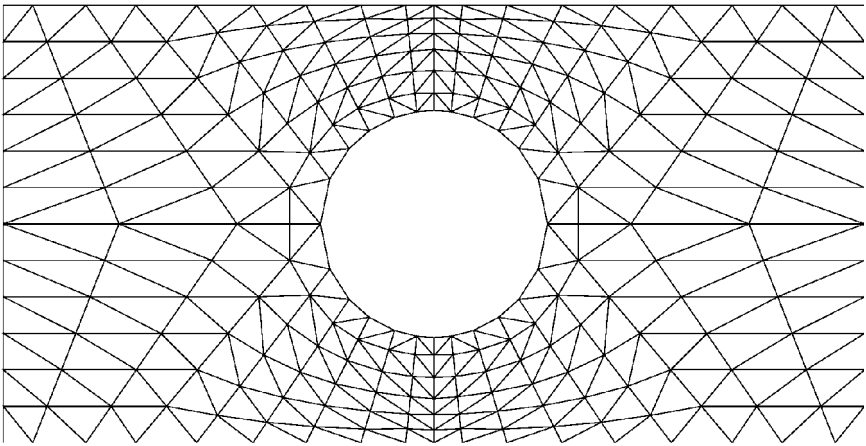


Figure 1.4 One important computer modelling technique is finite element analysis where the material part is simulated with a mesh. For each cell within the mesh, the physical variables are calculated and continuity over the interfaces is checked when the simulated material is stressed in some way.

node to node throughout the mesh. Likewise, changes in temperature, moisture content or any other material parameters can be simulated using this technique.

This modelling function is also enhanced by the computer's ability to recreate a virtual reality, 3D scene from sets of 2D image planes, and effortlessly produce special 2D cuts through the 3D data (as exemplified by confocal laser scanning microscopy in Chapter 4). An 'artistic' visualisation of the 3D structure of a glass fibre-reinforced composite, reconstructed using the confocal laser scanning microscopy technique, is shown in Fig. 1.5.

1.2 Characteristics of EM radiation

> And God said, let there be light, and there was light …
>
> … Genesis

1.2.1 Mathematical modelling and physical reality

Like the author of Genesis, a very pragmatic approach will be adopted in this book towards the central player in optical microscopy – namely 'light'. For many millennia, natural philosophers and scientists have been trying to explain the nature of light. Great strides have been made, but, even with the interest in String Theory (the latest candidate for a Grand Unified Theory of everything),[1] ultimately some leap of faith is required to accept the latest theoretical models seeking to explain the linkages between fundamental forces, elementary particles and light. Why? Because when we try to explain the nature and processes of sub-atomic space, we can only look for analogies with our everyday

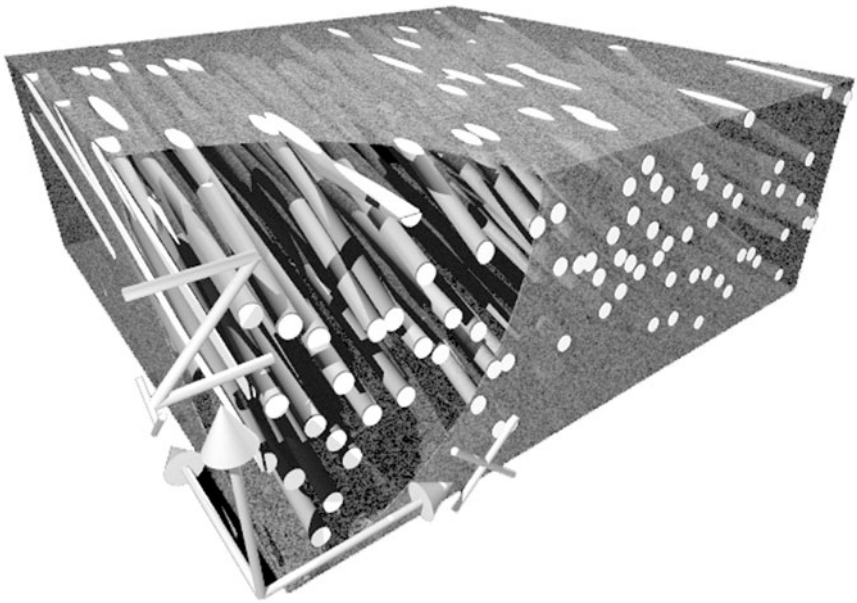


Figure 1.5 A typical 3D visualisation of glass fibres in a polymer composite, which has been derived from confocal laser scanning microscopy (CLSM) data – see Chapter 4. Realistic 3D simulations are now possible by applying appropriate shading paradigms like Phong shading.

experiences and, more importantly, the most sensitive measurement instruments cannot measure directly these underlying phenomena. Hence, we cannot provide direct experimental evidence to evaluate these new theoretical constructs (in the same way that we can evaluate the predictions of classical mechanics in everyday life or by reference to astronomical events).

Fortunately, the study of materials using microscopical methods only requires that material structure is explored down to the atomic, nanoscale sizes (10^{-9} m) rather than the incredibly small String Theory domain (10^{-20} m). For most purposes, especially research into mesoscale material structure, a spatial resolution of 100 nm will be more than adequate. In order to understand how most optical microscopes work and to use them effectively, one can approximate the 'true' nature of light propagation (whatever that might be) by using the simple analogies of classical wave motion and geometrical ray tracing. However, when describing the interactions of light with materials, it is necessary to suspend disbelief and treat light as a stream of discrete bundles of energy called 'photons'. Since the heated discussions between followers of Newton and the followers of Huygens in the seventeenth century, light has been regarded as either a particulate phenomenon or as a wavelike phenomenon, respectively. To the scientifically uninitiated, light must surely be one thing or the other but, to the pragmatist, its effects can be explained better in terms of

either particulates (photons) or waves depending on the circumstances. Indeed, the famous British scientist, Sir Arthur Eddington proposed the word 'wavicle' to describe this strange state of affairs. He also suggested that a useful comparison could be made between the wave/particle aspects of light and the more familiar act of coin tossing. When the coin is in the air, one does not know whether it is 'heads' or 'tails' until it hits the floor and is forced into one or the other state. In other words, the experimentalist should not worry about it being a wave or photon – until a measurement is made to determine its actual behaviour. The act of measurement forces it into behaving in one particular way.

As will become apparent, James Clerk Maxwell's electromagnetic theory gives a coherent explanation of the propagation of light, and the quantum theory developed by Bohr, Schroedinger, Heisenberg, Dirac and others describes the interaction of light and matter (the absorption and emission of light) very well. There are many books giving detailed expositions of electromagnetic theory, optics and quantum mechanics (see the bibliography at the end of this chapter). Therefore, for brevity and practicality, the approach taken here will be to present only those concepts, which are relevant to the microscopical techniques discussed later in the book.

1.2.2 Electromagnetic waves

A mathematical model of a wavelike motion is based upon our commonsense experience of typical waves, for example, ocean waves, ripples on ponds and vibrations on string instruments. These observations lead us to generalise wave motion, as shown in Fig. 1.6, where the travelling wave (a wave that transports energy from one place to another) is characterised by its amplitude, A, its wavelength, λ, its frequency, f, and its velocity, ν_s.

The general scalar wave equation describing the motion of a disturbance, U, through space is given by

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = \frac{1}{\nu_s^2}\frac{\partial^2 U}{\partial t^2} \qquad [1.1]$$

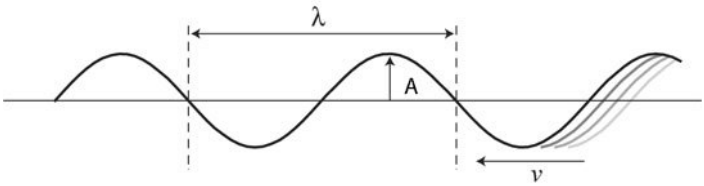


Figure 1.6 All harmonic waves in nature can be described by a mathematical model where the wave has amplitude, A, and a wavelength, λ. As the wave travels, the number of wave peaks passing a point in space per second is called the wave frequency, f, and the wave speed is given by the product $f\lambda$.

Consider the propagation of a wave in one dimension – say along the *z*-axis. This general equation reduces to

$$\frac{\partial^2 U}{\partial z^2} = \frac{1}{\nu_s^2}\frac{\partial^2 U}{\partial t^2} \qquad [1.2]$$

and it is straightforward to show that a solution to this equation is

$$U(z,t) = U_0 \cos(kz - \omega t) \qquad [1.3]$$

where k is the wavenumber $(2\pi/\lambda)$ and ω is the angular frequency $(2\pi f)$ of the wave motion. The term ωt is called the *phase* of the signal – in effect it determines the location of the maximum disturbance.

As the wave represents a flow of energy between two points in space, one must specify the type of wave motion. Is the host medium allowing transverse (perpendicular to the direction of motion) and/or longitudinal (parallel to the direction of motion) fluctuations to propagate? In the case of sound waves in air or liquids, the waves are longitudinal, i.e. alternate high and low pressure regions form due to compression or rarefaction of air molecules along the direction of propagation of the sound waves. However, sound waves in solids may be propagated by two different modes: transverse (or shear) waves and longitudinal (or tensile) waves, as illustrated schematically in Fig. 1.7.

For shear pressure waves in solids, if N is the shear modulus and ρ is the density of the solid, the velocity of the shear waves, ν_s, is given by

$$\nu_s = \sqrt{\frac{N}{\rho}} \qquad [1.4]$$

Different types of waves and their velocities in common materials are shown in Table 1.1.

Maxwell[2] was responsible for the major breakthrough when he established that light waves can be thought of as synchronous electric and magnetic field variations, which are perpendicular to each other and to the direction of

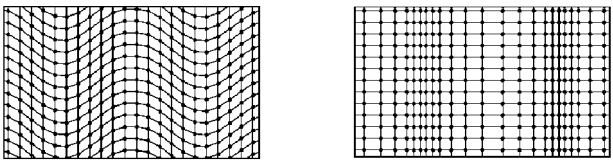


Figure 1.7 In order to visualise both transverse and longitudinal waves in materials, consider a crystal structure where all atoms are equally spaced in a rectangular array. (a) A transverse wave motion is shown where the wave moves to the right, but the atoms are moving up and down. (b) In this case, a longitudinal wave is shown where the atoms are oscillating about their mean position in the same direction as the wave motion, creating alternate dense and rarefied regions.

Table 1.1 Different types of waves and their velocities in materials

Wave type	Material	Velocity (metres/second)
Electromagnetic	Air	3×10^8
	Water	2.25×10^8
	Glass (typical)	2×10^8
Acoustic (longitudinal)	Dry air	331.5 (@ 293 K)
	Distilled water	1482.3 (@ 293 K)
	Crown glass	5660 (@ 293 K)
Acoustic (shear)	Crown glass	3420 (@ 293 K)
	Air and water	Not propagated in liquid or gas

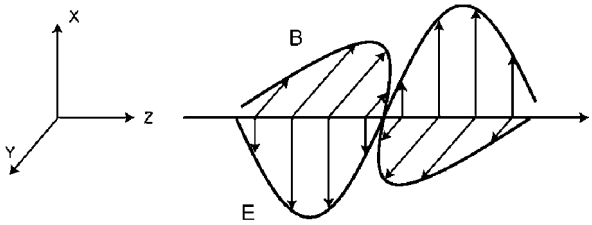


Figure 1.8 The wave view of electromagnetic radiation is that spatial variations in the magnetic B-field are synchronised, but perpendicular to variations in the electric E-field. The direction of travel is determined by the vector cross-product of E and B (Fleming's right-hand rule determines the direction of propagation).

propagation of the light waves (see Fig. 1.8). His famous equations for non-magnetic, non-conducting materials involving the magnetic field strength, H, the electric field strength, E, and a parameter called the displacement current, D, formed the basis of electromagnetic theory, as shown below.

$$curlE = -\frac{\partial B}{\partial t}$$

$$curlH = -\frac{\partial D}{\partial t}$$

$$divB = 0$$

$$divD = 0 \qquad [1.5]$$

where $D = \varepsilon E$, $B = \mu H$ and, because of the mathematical identity,

$$curl(curlf) = -\nabla^2 f + grad(divf) \qquad [1.6]$$

it can be shown that

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\varepsilon E) = \nabla^2 E \qquad [1.7]$$

which is another way of writing the wave equation, where the velocity of the electromagnetic wave, c, is

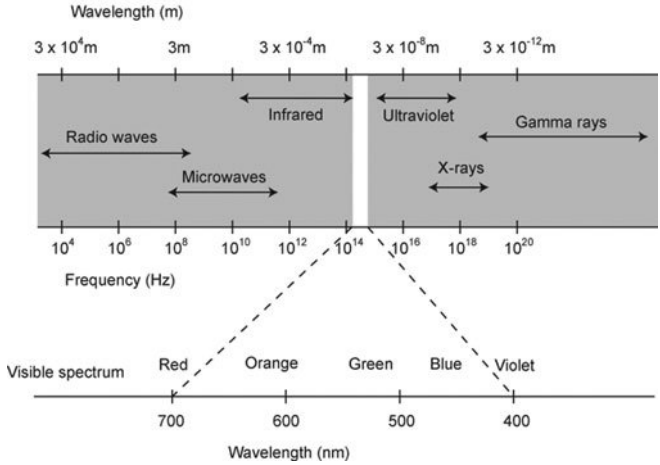


Figure 1.9 Light waves are a small part of the electromagnetic spectrum, which extends from long wavelength (low frequency) radio waves through to high frequency (small wavelength) gamma radiation. The sensation of colour depends on the wavelength of light.

$$c = \frac{1}{\sqrt{\mu\varepsilon}} \qquad [1.8]$$

The physical parameters which determine the velocity of light are μ (the permeability of the material) and ε (the permittivity of the material). The triumph of Maxwell's theory was that it not only described optical phenomena but also explained the link between light and other forms of electromagnetic phenomena like x-rays and radio waves, the only difference between these disparate phenomena being the wavelength of the electromagnetic radiation. The wavelength ranges and their different classifications, which constitute the electromagnetic (EM) spectrum, are shown in Fig. 1.9.

Sources of electromagnetic radiation will be considered in more detail later, but note that there are two types of source: *coherent* and *incoherent*. EM radiation is formed by charges within the source oscillating and, if the charges oscillate in unison with each other, the resulting radiation emitted will be coherent (as exemplified by the laser). However, in all other laboratory sources, the optical radiation is produced by charges oscillating independently and randomly and hence these sources produce incoherent light.

1.2.3 Refraction and refractive index

These EM waves may be thought of as distortions in an electromagnetic field which pervades space and the in-phase electric and magnetic field variations have an amplitude, A. The waves have a wavelength λ (in metres) and frequency f (in Hertz), as shown in Fig. 1.8. The eye perceives brightness, which is related

Table 1.2 Refractive indices of materials (at wavelength of 550 nm)

Material		Refractive index (at $\lambda = 589$ nm)	Comments
Air		1.000292	STP conditions ($\theta = 293$ K)
Water		1.3330	$\theta = 293$ K
PTFE		1.325	
PA (nylon 66)		1.448	
PVC		1.542	Refractive index can vary widely
Medium Barium Crown Glass		1.5694	A typical glass
Quartz	(O-ray)	1.544	$\theta = 293$ K
	(E-ray)	1.539	
Calcite	(O-ray)	1.660	$\theta = 291$ K
	(E-ray)	1.486	
Diamond		2.4195	

to the intensity, I, of these waves (where I is proportional to A^2), and colour, which is directly related to the wavelength, λ. The velocity of propagation of the light waves is given by c (which is the product of the frequency and wavelength, $c = \nu\lambda$) and the velocity of light is dependent on the material through which it is travelling. The fastest velocity, $c_o = 3.10^8$ m/sec is through empty space (a vacuum) and the velocity through a material, c, is determined by that material's refractive index, n.

$$n = c_o/c \quad \text{or} \quad c = c_o/n \qquad [1.9]$$

Hence, the refractive index of a vacuum is unity and air and other gases have a refractive index very close to unity. Values of the refractive indices, n, of some common materials are shown in Table 1.2.

When describing optical effects, which are linked to the relative phases of light waves travelling over different paths within a material, the optical pathlength difference, $n\delta l$, rather than the physical pathlength difference, δl, must be computed. Assuming that two waves were in phase before travelling over different paths, the phase difference between them after δl will be

$$\Theta = \frac{2\pi}{\lambda} n\delta l \quad \text{where } \Theta \text{ is in radians.} \qquad [1.10]$$

Also, the refractive index determines the angle through which a planar wavefront incident on an interface between two materials will be refracted or bent from its original direction of motion. In general, a light ray incident on an interface at an angle θ_1 will split into a reflected ray and a transmitted ray (see Fig. 1.10(a)) whose direction to the normal at the interface θ_2 is given by Snell's Law

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \qquad [1.11]$$

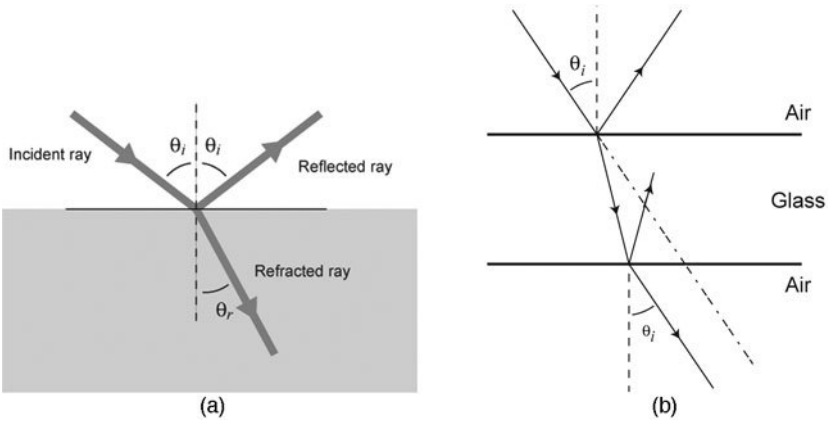


Figure 1.10 (a) When light strikes an optically flat surface of a high refractive index material (say glass) in a lower refractive index material (say air), part of the energy will be reflected and the rest will be transmitted through the interface region. The refracted ray makes an angle, θ_r, to the normal, where $n_g \sin \theta_r = n_a \sin \theta_i$ and the refractive indices of glass and air are n_g and n_a respectively. (b) When the light ray passes through a parallel-sided glass sheet, the emergent ray will be travelling in the same direction as the original light ray but it will be translated sideways by a small distance.

Note this equation implies that

(a) light at normal incidence ($\theta_1 = 0^o$) to an interface will be transmitted without deviation from its original path (but with some attenuation in intensity)

(b) if the light ray is passing from a medium of lower refractive index, n_1, to a medium of higher refractive index, n_2, the light ray will be refracted (deviated from its original direction) and some fraction of the light will be reflected too. When the light beam passes through a parallel-sided sample, the light beam will exit the sample travelling in the same direction as the incident beam but translated sideways (see Fig. 1.10(b)).

(c) if light passes from a medium of higher refractive index, n_2, to lower refractive index, n_1, the angle of refraction will be greater than the angle of incidence, until, for a critical angle of incidence, θ_c, the angle of refraction becomes 90°. For higher angles of incidence, the light will be *totally internally reflected* and no light will be transmitted through the interface.

$$\theta_{crit} = \sin^{-1}\left(\frac{n_1}{n_2}\right) \qquad [1.12]$$

For example, when light passes from glass ($n_2 = 1.5$) into air ($n_1 = 1$), the critical angle is $\theta_{crit} = 41^o$.

This phenomenon has been utilised successfully in the fabrication of optical fibres where an inner glass core has a refractive index, n_{inner}, which is lower than

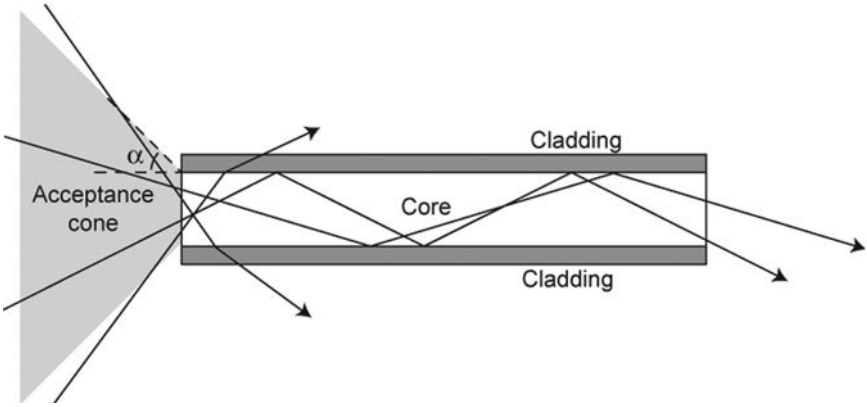


Figure 1.11 Optical fibres consist of an inner core of one refractive index, surrounded by a cladding of higher refractive index. It can be shown that all light rays within an acceptance angle, α, will travel along the inner core region of the optical fibre over long distances with little attenuation.

the surrounding outer sheath refractive index, n_{outer}. The acceptance cone of rays (within angle α) entering into an optical fibre and which will be totally internally reflected by the fibre core, is given by

$$\alpha = \sin^{-1}\sqrt{(n_{inner}^2 - n_{outer}^2)} \qquad [1.13]$$

As shown in Fig. 1.11, the light rays are trapped within the fibre and hence they can be transported with little attenuation over large distances and into confined regions of space (e.g. endoscopes for key-hole surgery). The main use of optical fibres in microscopy is for directing intense illumination onto the specimens in a very flexible way but, recently, special confocal microscope designs have also incorporated the direct use of optical fibres in the imaging process.[3] Note, however, that the refractive index values are functions of wavelength and can vary markedly, especially near to absorption bands, as shown in Fig. 1.12(a). In Fig. 1.12(b), the refractive index variation of different types of glass used in microscopy is shown across the visible region and into the ultraviolet.

1.2.4 Group velocity

So far, the electromagnetic wave has been regarded as a plane harmonic wave with a unique wavenumber and angular frequency. However, if one considers the superposition of two waves with the same amplitude but slightly different angular frequencies $(\omega + \Delta\omega)$ and $(\omega - \Delta\omega)$ and correspondingly different wavenumbers $(k + \Delta k)$ and $(k - \Delta k)$, the resulting wave will be described by the summation of two continuous waves

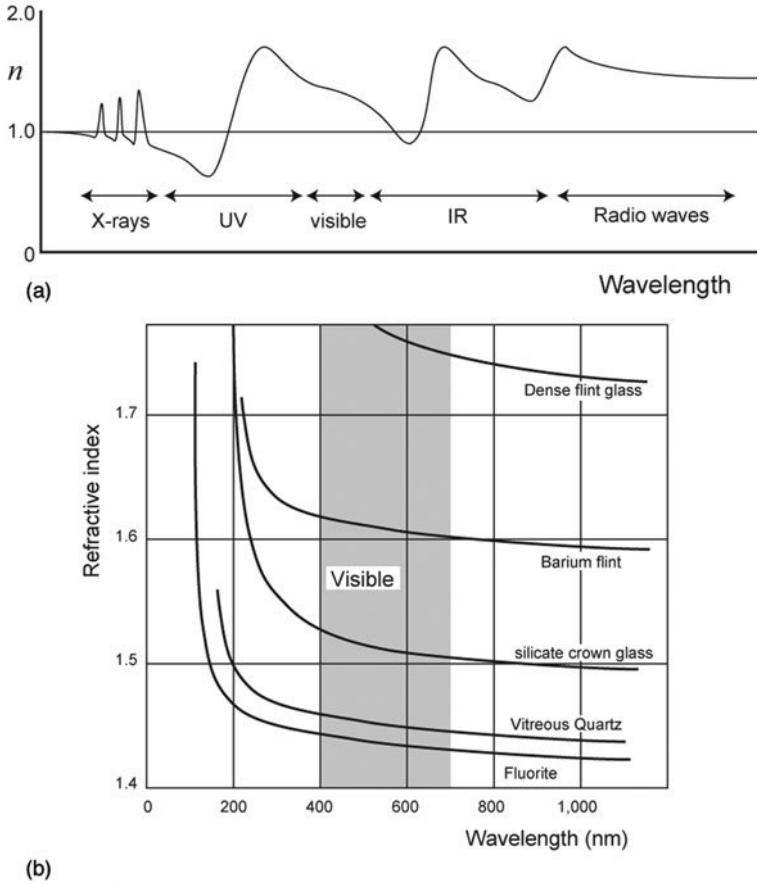


Figure 1.12 (a) Schematic diagram of the complete dispersion curve for a material which is transparent to the visible part of the EM spectrum. When $dn/d\lambda$ is positive, the material absorbs strongly. (b) More detailed dispersion curves for some materials often used for lenses and prisms are shown. Note that $dn/d\lambda$ is negative over the visible spectrum (and therefore, dn/dk will be positive in this region).

$$U = U_{k+\Delta k} + U_{k-\Delta k} \qquad [1.14]$$

which, after manipulation, gives an equation of the form

$$U = 2U_0 e^{\beta} \cos(z\Delta k - t\Delta\omega) \qquad [1.15]$$

The interpretation of this equation is that a single wave motion travels along the z-axis as shown in Fig. 1.13. Rather than a continuous cosine wave being propagated, a modulation envelope is propagated. If, instead of the two waves shown in Fig. 1.13, a larger number of similar waves (of angular frequencies within the range $\Delta\omega$ and range of wavenumbers Δk) are superposed, a single pulse or finite modulation envelope, rather than the multiple beats shown in Fig.

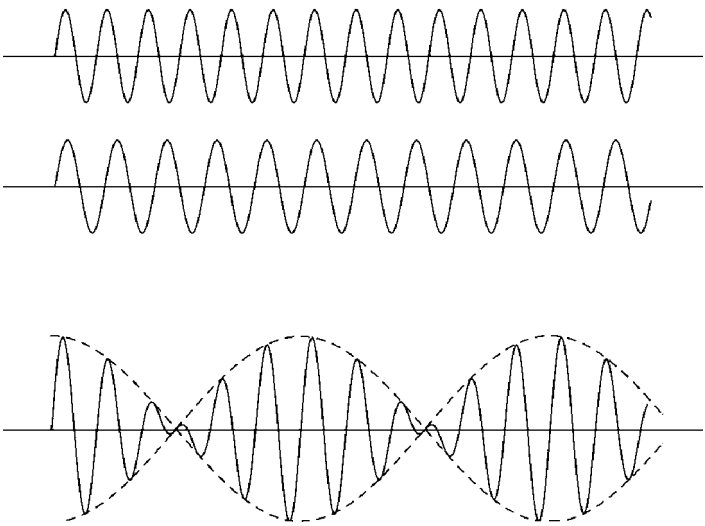


Figure 1.13 When two waves of the same amplitude but slightly different frequencies combine in the same region of space, they create the amplitude modulated pattern shown. The variations in amplitude travel through space with a group velocity (rather than the velocity of either wave). If a signal is composed of a finite range of similar frequencies, which combine in this way, instead of a large number of amplitude modulated patterns, a short duration pulse will propagate at the group velocity – as discussed later in Chapter 6.

1.13, would propagate. The modulation envelope would have a velocity, ν_g, given by

$$\nu_g = \frac{\Delta\omega}{\Delta_k} \qquad [1.16]$$

and ν_g is referred to as the group velocity (as distinct from the phase velocity, v_s). In all optical media, the phase velocity of the EM radiation, v_s, is a function of the angular frequency, ω. This is the phenomenon of *dispersion*. For most optical media, the refractive index varies as the frequency or wavenumber in such a way that *dn/dk* is positive (see, for example, Fig. 1.12(b)). As it can be shown that

$$\nu_g = \nu_s\left(1 - \frac{k}{n}\frac{\partial n}{\partial k}\right) \qquad [1.17]$$

it follows that the group velocity will be less than the phase velocity when *dn/dk* is positive. Any signal can be considered to be a modulation of a carrier wave. The signal or information content in the wave (or specific disturbance) propagates with the group velocity, v_g, rather than the phase velocity. In the case of pulses of nearly monochromatic light, i.e. with a small spread of wavelengths around a central wavelength, this group velocity is well defined.

In order to determine the propagation of EM energy per unit area, the average value of the Poynting vector, **S**, may be computed

$$\mathbf{S} = \mathbf{E} \wedge \mathbf{H} \qquad [1.18]$$

where the symbol '$\wedge$' denotes the vector cross-product and **S** is expressed in watts per m^2. The average Poynting flux is $< S >= I.\mathbf{n}_z$ where I is the irradiance and $\mathbf{n}_z$ is the direction of propagation of the electromagnetic wave. In isotropic media, the direction of energy flow, $\mathbf{n}_z$ is in the same direction as the wavevector, **k** (whose magnitude is the wavenumber) as one might expect. However, in non-isotropic media like crystals, **S** and **k** are not always acting in the same direction.

1.2.5 Polarisation states

Light waves can be regarded as transverse electromagnetic waves and another important parameter characterising these waves is their *polarisation* state. Polarisation is historically specified by reference to the electric field orientation. If the electric field variations remain parallel to a fixed plane in space, the wave is linearly polarised, as shown in Fig. 1.14(a). The polarisation of a plane-polarised light wave may, therefore, be represented by a vector whose orientation is fixed along the path of the light wave. In the case of natural or unpolarised light, the instantaneous polarisation fluctuates rapidly in a random manner (any vibration plane orthogonal to the direction of travel being equally likely). Basically, there are four phenomena that can produce polarised light from unpolarised light: absorption, scattering, reflection and birefringence (also called double refraction). By studying the polarisation state of light waves, it is possible to obtain valuable information about these physical phenomena. Hence

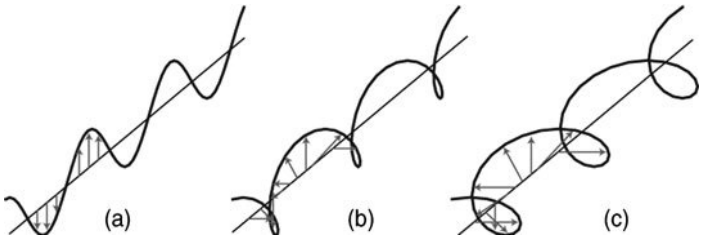


Figure 1.14 Different types of wave polarisation phenomena are shown schematically. Consider the resultant E-field vector only. As the wave propagates, in (a) *plane polarisation*, the E-field stays in one fixed direction, while in (b) *elliptical polarisation,* it rotates as it travels and its amplitude varies too, and in (c) *circular polarisation,* it rotates but its amplitude stays constant. The resultant vector for circular and elliptical polarisation could rotate clockwise or anti-clockwise, from the perspective of an observer receiving the light waves, and the light is said to be right-handed or left-handed respectively.

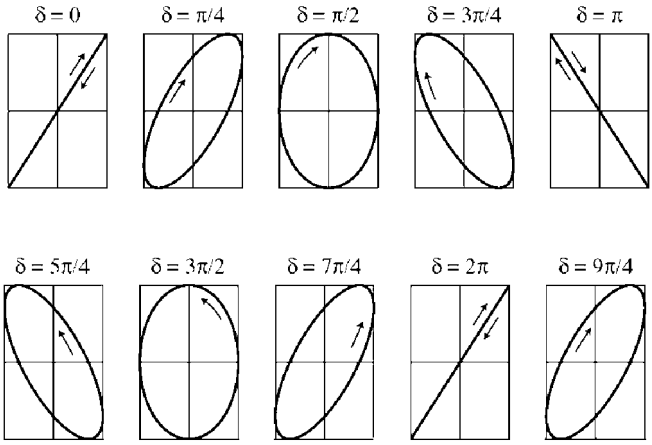


Figure 1.15 We cannot 'see' the polarisation states of light waves, but an interesting experimental analogue is the Lissajous figure seen on an oscilloscope screen when two waves of the same frequency are combined orthogonally (by ensuring that one signal controls the *x*-deflection and the other signal the *y*-deflection on the oscilloscope). If the phase difference between these signals is given by δ, it can be seen that $\delta = 0$ or π (or 2π) radians corresponds to the linear polarisation state. The other phase combinations look like elliptical polarisation states – but if the amplitudes of the combined signals had been equal, the $\delta = \pi/2$ or $3\pi/2$ would correspond to right-handed or left-handed circular polarisation states, respectively.

the importance of polarisation microscopy where these material processes in the specimen can be investigated using polarising elements within microscope systems.

There are a number of different types of polarisation state besides linear polarisation and unpolarised light and they can be understood by analogy to Lissajous figures in electronics. Lissajous figures are generated on an oscilloscope screen when two sine waves with the same frequency (or harmonically related) are input to channel 1 and channel 2 of the oscilloscope and the timebase of the oscilloscope is turned off, as shown in Fig. 1.15. The two sine waves, one representing variations in the *x* direction and the other representing variations in the *y* direction, are generally out of phase with each other. If the waves have the same frequency, the Lissajous figure is static (the shape does not alter) but if the waves have different frequencies, a time-varying figure would be seen. Depending on the relative amplitudes of the two waves, the resulting static figure of the combined *x* and *y* variations is given by summing the following two expressions with phases α_1 and α_2,

$$y = a_1 \sin(\omega t - \alpha_1)$$
$$x = a_2 \sin(\omega t - \alpha_2) \qquad [1.19]$$

After some manipulation, it can be shown that

$$\sin^2(\alpha_1 - \alpha_2) = \frac{y^2}{\alpha_1} + \frac{x^2}{\alpha_2} - \frac{2yx}{\alpha_1\alpha_2}\cos(\alpha_1 - \alpha_2) \quad [1.20]$$

describes the relationship between x and y and hence the shape of the figure. The phase difference between the two light waves is given by

$$\delta = \alpha_1 - \alpha_2 \quad [1.21]$$

When $\delta = 0, 2\pi, 4\pi, \ldots$ the equation reduces to

$$y = \frac{a_1}{a_2}x \quad [1.22]$$

and when $\delta = \pi, 3\pi, 5\pi \ldots$, the equation reduces to

$$y = -\frac{a_1}{a_2}x \quad [1.23]$$

Linear or plane-polarised light waves could be represented by one of these sine waves, which are varying in only one plane. Therefore, for the sake of argument, imagine that the electric field strength displacements are linear displacements in the $y-z$ plane (and that the wave is travelling in the z-direction). If this wave were combined with another linear or plane-polarised light wave, where the displacements are restricted to the $x-z$ plane, the resultant motion of the tip of the electric field strength wave vector would draw out an ellipse in x and y values. Hence, this is called elliptically polarised light, as shown in Fig. 1.14(b).

If the amplitudes of the two light waves are equal ($a_1 = a_2$) and the phase difference is an odd multiple of 90º, the tip of the electric field strength vector executes a circular motion in the x and y values and is called circularly-polarised light. Furthermore, if the direction of rotation of the displacement in the $x-y$ plane is clockwise (when looking opposite to the direction of propagation), it is called right circularly-polarised. However, if the rotation is counter-clockwise and the phase difference is an odd multiple of 90º, it is called left circularly-polarised light. Circular polarisation is illustrated in Fig. 1.14(c).

1.2.6 Polarisers and quarter-wave plates

An important polarising element often found in microscopy is the linear polariser. In 1938, Edwin Land invented a special film (called Polaroid film) of oriented long-chain hydrocarbon molecules which became conducting when the film had been dipped in iodine solution (and he also invented the Polaroid film processing procedure). When the electric field vector is parallel to these chains, electric currents are set up along the chains and the light energy is absorbed. If the electric field is perpendicular to the chains, the light is transmitted with little attenuation. Such a film is called a polariser and the direction perpendicular to

the chains is called the transmission axis. Malus' Law relates the output intensity, I, of the linearly polarised light wave to the initial intensity I_0 (when it passes through a polariser whose transmission axis is at an angle θ to the plane of polarisation of the incident wave).

$$I = I_0 \cos^2\theta \qquad [1.24]$$

The quarter-wave plate is a device for creating circular polarisation from two linearly polarised beams. Some transparent crystals like calcite or mica are 'doubly-refracting' (or birefringent) because their refractive index has two different values for two different directions of polarisation of an incoming light beam, as shown in Fig. 1.16(a). The quarter-wave plate is a specially cut crystal slab of thickness d, such that it has a slow axis (maximum index of refraction, n_1) at right angles to a fast axis (minimum index of refraction, n_2). The thickness is chosen to make the optical path difference between two orthogonal, linearly polarised beams equal to one quarter wavelength at the incident light wavelength, λ_0.

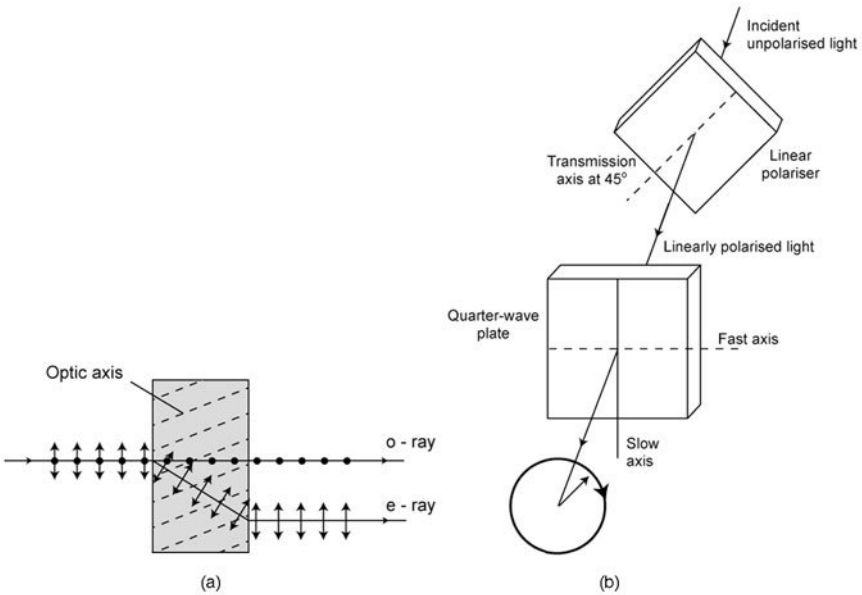


Figure 1.16 (a) Some crystals exhibit birefringence. Consider unpolarised light incident normally on such a crystal. On entering the crystal, a linearly polarised, ordinary ray (o-ray) will propagate normal to the interface and an extra-ordinary ray (e-ray), which is linearly polarised at right angles to the o-ray, will be refracted as shown. (b) This effect may be used to create circularly polarised light from unpolarised light by combining a linear polariser before a quarter wave plate constructed from a birefringent material.

$$d = \frac{\lambda_0}{4(n_1 - n_2)} \quad [1.25]$$

Therefore, if unpolarised light is incident on a linear polariser whose transmission axis is at 45° to the slow axis of the quarter-wave plate, two in-phase, linearly polarised beams will be incident on the quarter-wave plate. However, they will be 90° out of phase when they exit the quarter-wave plate and hence will display right circular polarisation, as shown in Fig. 1.16(b). If the linear polariser is rotated by 90°, it can be shown that a left circularly-polarised light beam will be produced by the quarter-wave plate.

1.3 Propagation of light waves

When considering the propagation of light waves, it is best to describe the general principles in terms of rays rather than considering the behaviour of the electric and magnetic field strength vectors. The ray merely represents the direction in which a wave travels. This simplification is the result of Fermat's Principle which states that the path of a light ray from one point to another is either a maximum or a minimum optical length (actual physical length multiplied by the refractive index). However, when quantitative predictions have to be made, Maxwell's equations must be invoked.

As discussed later, to interpret microscopic images and appreciate the practicalities of making measurements, materials are probed with waves (light waves or ultrasonic or x-rays) and material structure is determined by how some property of the wave is affected during its transmission through or reflection from the material. The fundamental property of the wave that determines the spatial resolution attainable is the wavelength, λ, and the wavelength is linked to two other wave properties: frequency, f, and velocity, c, by the simple equation

$$\lambda = c/f \quad [1.26]$$

However, in order to record an image there must be spatial variations in light intensity (in the square of the electric field strength) over the field of view. Maxwell's equations may be applied (with suitable boundary conditions to ensure, for example, the conservation of energy) at material interfaces in order to predict the reflection from and the transmission through the interface. In other words, the reflection coefficient, R (ratio of reflected and incident intensities), and the transmission coefficient, T (ratio of transmitted and incident intensities), may be predicted. All that is needed is the polarisation state of the incident light and also the respective refractive indices on either side of the interface.

The first topic to consider is the specular reflection of light from flat surfaces. The most fundamental relationship in optics is $\theta_i = \theta_{ref}$ where the angle of incidence is θ_i and the angle of reflection is θ_{ref}. However, as shown in Fig. 1.17(b), this relationship only holds for optically flat surfaces (which means that any residual roughness of the surface has a scale size of the order of $\lambda/10$ or

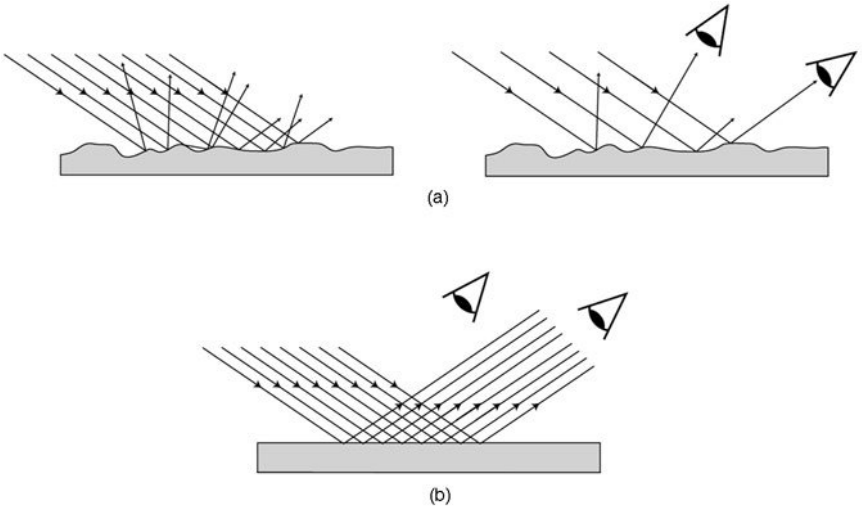


Figure 1.17 (a) When the surface of a reflecting material has surface height variations comparable with, or larger than, the wavelength of the incident light, λ, incident light rays will be reflected in all directions and hence the surface appears dull or matt to an observer. (b) When height variations are less than $\lambda/10$, the surface is said to be optically flat and the surface exhibits specular reflection or mirror-like performance.

better, where λ is the wavelength of the light being reflected). If the surface is not optically flat, the light will be diffusely reflected, as shown in Fig. 1.17(a) and the reflected light will have a wide angular dependency, making the surface appear dull and matt.

1.3.1 Reflection of light at flat surfaces and interfaces

Considering light as the flow of energy from one point in space to another, then we have to believe that, at any point in space, there is conservation of energy. Hence, when light meets an interface created by two dissimilar materials (like air and glass), we would expect that the light is partially transmitted and partially reflected. If the reflectivity or fraction of reflected light intensity is R and the fraction of transmitted light intensity is T, for ideal metallic mirrors, $R = 1$ and $T = 0$. One might think that when light is incident on a transparent glass surface, the opposite situation occurs, i.e. $R = 0$ and $T = 1$. However, when light falls at normal incidence on an air–glass (or glass–air) interface, a small amount of the light is reflected. The fraction of reflected light intensity, R, may be deduced from Maxwell's equations and is given by

$$R = \left[\frac{(n_1 - n_2)}{(n_1 + n_2)}\right]^2 \qquad [1.27]$$

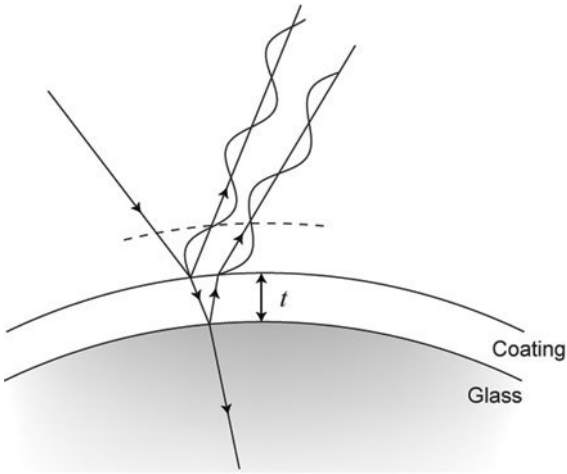


Figure 1.18 The blooming or coating of a lens surface with a thin film of material whose refractive index, n_b, differs from that of glass can lead to destructive interference between the ray reflected off the surface of the coating and the ray reflected off the glass surface, as shown. Hence, the (spurious) reflectivity of the lens is minimised.

where n_1 and n_2 are the refractive indices of the two transparent media on either side of the interface. Optical glasses (with refractive indices close to 1.5) therefore reflect around four per cent of the incident light at the air–glass interface. This may seem a small fraction, but if the light traverses, say m separate air–glass interfaces, the amount of light lost (or giving annoying reflection artefacts within an optical system) becomes $(4/100)^m$. In the twentieth century, this potentially problematic effect has been improved by the technique of 'blooming' the lens, i.e. applying a thin coating, thickness t, of a material of refractive index, n_b, onto the glass surface, as shown in Fig. 1.18. The rationale of this technique is to make the two reflected rays at the air–bloom material and the bloom–glass interface destructively interfere at a wavelength around 550 nm. For interference to take place, two conditions must be met

1. The reflected intensities at each interface must be the same, therefore

$$\left[\frac{(1-n_b)}{(1+n_b)}\right]^2 = \left[\frac{(n_b-n_g)}{(n_b+n_g)}\right]^2 \qquad [1.28]$$

and after some mathematical manipulation, this implies that

$$n_b = \sqrt{n_g} \qquad [1.29]$$

2. The extra optical pathlength within the bloom material is an odd multiple of half wavelengths, implying that the thinnest bloom layer should have a thickness

$$t = \frac{\lambda}{4n_b} \quad [1.30]$$

These coatings may be evaporated onto the glass surfaces under vacuum and the lenses usually appear purple or magenta in colour. In practice, lenses which have been bloomed have multiple coatings (each with a different refractive index), rather than a single coating.

1.3.2 Angular dependence of the specular reflectivity

As the electric and magnetic fields are orthogonal to each other when linearly polarised light is incident on an interface, there are two different ways that the light can interact with the surface. Either the electric field is parallel (i.e. transverse) to the surface plane (called the TE mode) or the magnetic field is parallel to the surface plane (called the TM mode), as shown in Fig. 1.19. By satisfying the boundary conditions at the interface and deriving the variation of reflectivity with incident angle, Maxwell's equations predict that there will be a certain incident angle – the Brewster angle, θ_B – with interesting properties. When unpolarised light is incident on the interface at this angle, the reflected waves become linearly polarised (with the electric field parallel to surface) and the transmitted waves become partially polarised, as illustrated in Fig. 1.20.

It can be shown that the Brewster angles are given by the equations

$$\theta_B = \tan^{-1}\left(\frac{n_2}{n_1}\right) \text{ for the air/glass interface, } \theta_B = 57^{\circ} \quad [1.31]$$

$$\theta_B = \tan^{-1}\left(\frac{n_1}{n_2}\right) \text{ for the glass/air interface, } \theta_B = 33^{\circ} \quad [1.32]$$

Maxwell's equations also predict the variation of reflectivity, R, with incident angle, θ_i, and, in the case of an air–glass boundary, this is illustrated in Fig. 1.21. Another key component within microscope systems is the beamsplitter, which is designed to partially reflect and partially transmit incident light. A glass

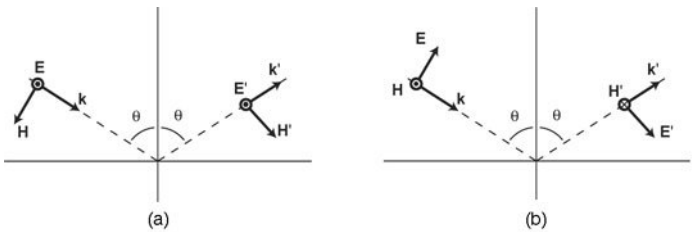


Figure 1.19 Maxwell's Electromagnetic Theory takes into account the variation in the reflection of EM waves when the E-field vector is parallel to the interface, as in (a) (TE mode) and when it is in the plane perpendicular to the interface, as in (b) (TM mode).

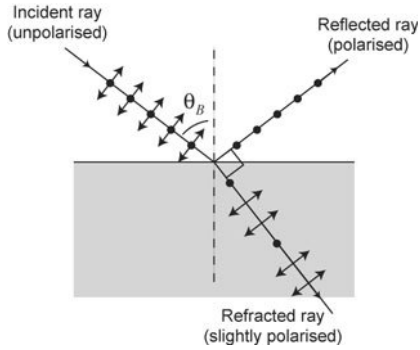


Figure 1.20 One curious reflection phenomenon, which was first established by Brewster, is the Brewster's angle condition. When an unpolarised light beam is incident on an interface, there is one angle of incidence for which the reflected beam is linearly polarised parallel to the surface and the refracted ray is partially polarised orthogonal to the reflected beam. (Note that, at this angle, the reflected and refracted beams are travelling at right angles to each other.)

substrate is half-silvered with a metal coating to create an optical component where $R = T = 0.5$. When it is placed at 45° to the incident beam, the transmitted light is propagated through the glass substrate and continues along essentially the same path as the incident beam. The reflected beam is propagated in a direction perpendicular to the incident light beam path. Hence, using either beamsplitters, prisms, optical fibres or flat mirrors, light rays may be manipulated in space, rather like electrical signals are routed by copper wires in computer systems.

1.3.3 Reflection off non-planar surfaces

So far, only planar surfaces have been considered which modify the direction of travel of the incident light beam and produce one or more images (either inverted or upright or rotated) but always of the same size as the object. However, if the light beam is reflected off a curved surface, it is possible to create smaller or larger images of an object. Indeed, it is possible to create a reflecting telescope (see Newton's design[4]) or a reflecting microscope (see Section 1.4.5). The imaging properties of curved reflecting surfaces are shown in Fig. 1.22. Rays, which are normal to the curved surface, will reflect back along their original path. Paraxial rays (rays parallel to the optical axis), when striking a spherical or parabolic reflecting mirror, are brought to the same point on the optical axis – the focus. Hence the position of the image of an object can be determined from the intersection of these rays. Therefore, if one considers the light rays from an object which is close to this focus point but on either side of the focus, the resulting images will be dramatically different, as shown in Fig. 1.22(a) and (b). Note also that when a bright light source is placed at the focus of

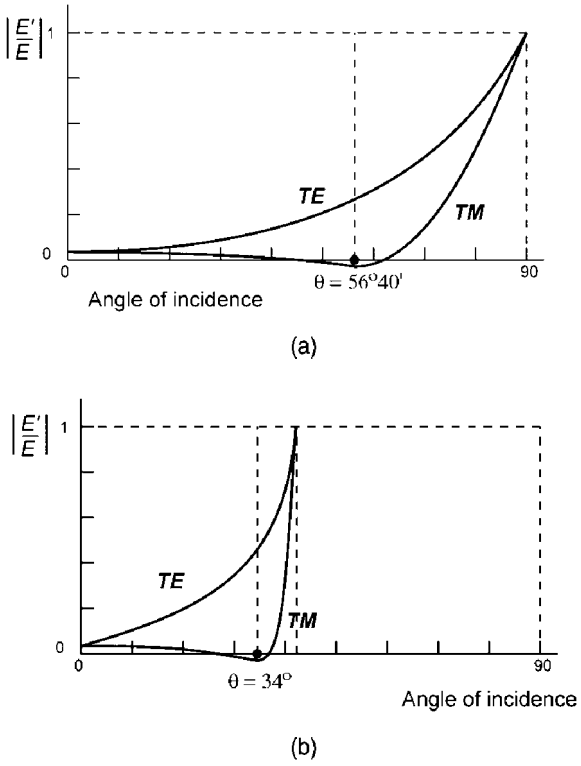


Figure 1.21 By solving Maxwell's equations, the reflectivity (ratio of reflected and incident amplitudes) can be expressed as a function of the incident angle. The variation in reflectivity for both TE and TM modes is shown for (a) external reflection (e.g. air to glass), and (b) internal reflection (e.g. glass to air). Note that, for internal reflection after an incident angle of 42°, the light rays are totally internally reflected.

a parabolic mirror, a collimated beam of light is produced, as in a searchlight, because the image is at infinity.

1.3.4 Interference between waves

Usually, when two or more incoherent sources of light (for example, electric light bulbs with intensities I_1, I_2, I_3) are brought close together, their intensities simply add together ($I_{total} = I_1 + I_2 + I_3$). Hence the overall illumination (the irradiance) at a point in space, a distance r from these sources, will be given by

$$\text{Irradiance} = I_{total}/4\pi r^2 \text{ watts/m}^2 \qquad [1.33]$$

on the assumption that the sources emit radiation isotropically.

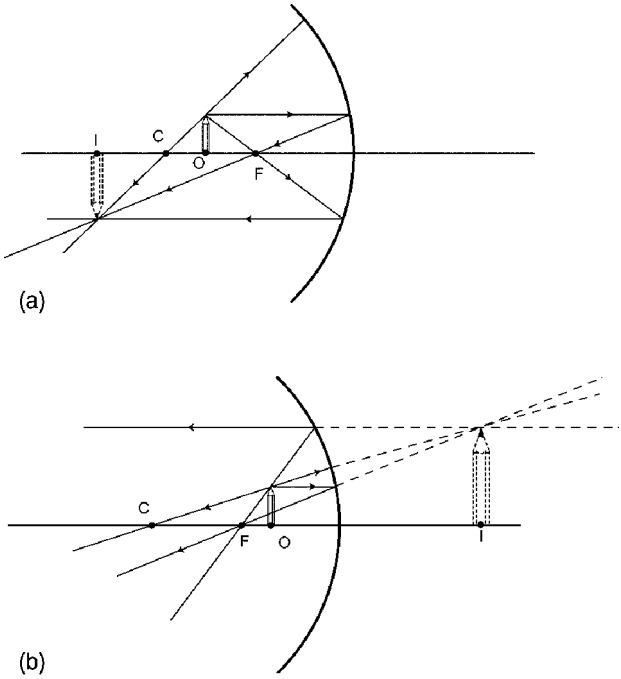


Figure 1.22 Ray tracing can be used to determine the size and location of an image. Consider the images formed by a concave spherical mirror surface when the object is placed (a) between the centre of curvature and focus and (b) between the focus and the mirror surface. The image is formed by selecting the ray through the centre of curvature (which strikes the mirror surface normally) and the rays through the focus (which, on striking the mirror surface, travel parallel to the axis).

These typical laboratory sources are said to be 'incoherent' because the billions of light bursts, which constitute the individual light intensities from each source, are varying randomly in phase. The overall intensity is therefore simply the average over all of these light bursts. However, if a means can be found (either by division of wavefront or amplitude) to split up one light beam into two beams, the two beams will become 'coherent'. At any instant in time, at the point of their creation, the light from one beam will always have a fixed phase in relation to the light from the other beam. Also, if the two beams are made to traverse the same optical pathlength before recombining, they will retain this phase difference with each other. When the beams recombine in another region of space, there will be a fixed phase difference between the beams at all times (see Fig. 1.23).

The relationship between the physical path difference, δl, between the beams and their relative phase difference, Θ, when propagating through a medium of refractive index, n, is given by

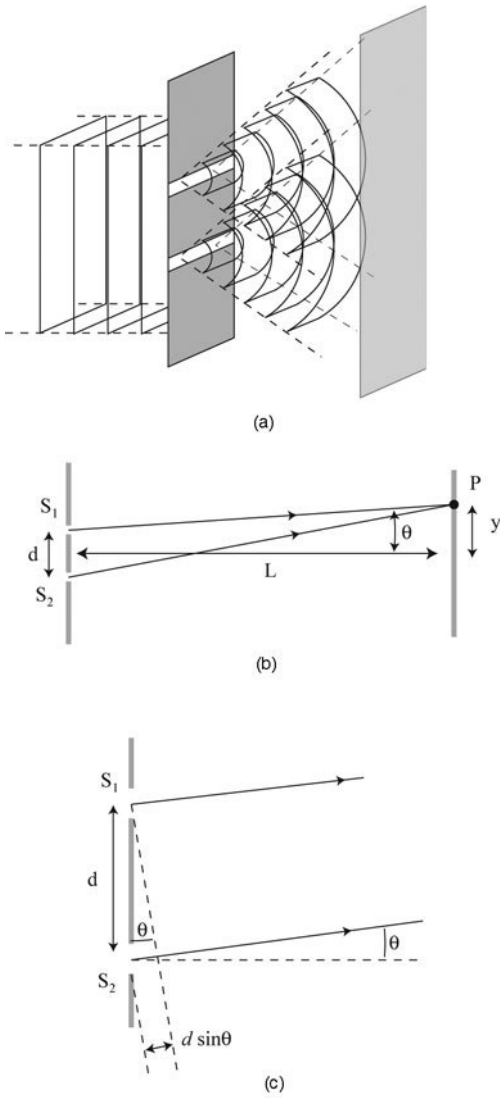


Figure 1.23 The basis of the interference effect is illustrated. (a) A plane wave incident on the screen, which has two narrow slits, will create two expanding wavefronts. These wavefronts will be coherent (in synchronisation with each other) and, where they overlap on the screen, they will constructively interfere (to produce light bands) or destructively interfere (giving dark bands) parallel to the slits. (b) In effect, two sources, S_1 and S_2 have been created which will constructively interfere at point P on the screen, if the optical path difference between the light rays is a multiple of the wavelength, λ, of the light. (c) As the screen distance L is far greater than the slit separation, the two interfering rays are essentially parallel and, hence, the optical path difference in air is given by $d \sin\theta$ where d is the separation of the slits and $\theta = y/L$.

$$\Theta = \frac{2\pi}{\lambda} n\delta l \quad \text{where } \Theta \text{ is in radians.} \qquad [1.34]$$

The classical illustration of this effect is the Thomas Young's slit experiment where a monochromatic source (producing light of a single wavelength, λ) after striking a screen with two parallel thin slits, creates two coherent light beams through division of wavefront. If the observer looks at the resulting light distribution on a viewing screen some distance away, alternate light and dark bands are seen. The high intensity bands are where the two waves are in phase with each other so that they constructively interfere and the dark bands are where they are 180° out of phase with each other so that they destructively interfere. Assume that the intensities of the two coherent beams are I_1 and I_2. Whenever the beams constructively interfere, the combined intensity will be $I_1 + I_2$ and when the beams destructively interfere, the resultant intensity will be $|I_1 - I_2|$ (see Fig. 1.24). The mathematical expression which relates the intensity at a point P on the viewing screen to the separation of two slits, d, and the angle off-axis, θ, is given by

$$I = 4I_0 \cos^2 \frac{\delta}{2} \quad \text{where } \delta = \frac{2\pi}{\lambda} d \sin\theta \qquad [1.35]$$

An example of interference by division of amplitude is in the Michelson interferometer which employs a beamsplitter (a parallel-sided glass plate with a lightly-silvered rear surface which allows a fraction of the incident light to be transmitted and a similar fraction to be reflected). These beams, after traversing different optical paths, as shown in Fig. 1.25, recombine to form interference fringes. Interferometers have had an important role to play in many fundamental physics research fields, for example, the same interferometry concept has been adapted for a number of sensitive instruments, which are currently searching for the existence of gravity waves.[5]

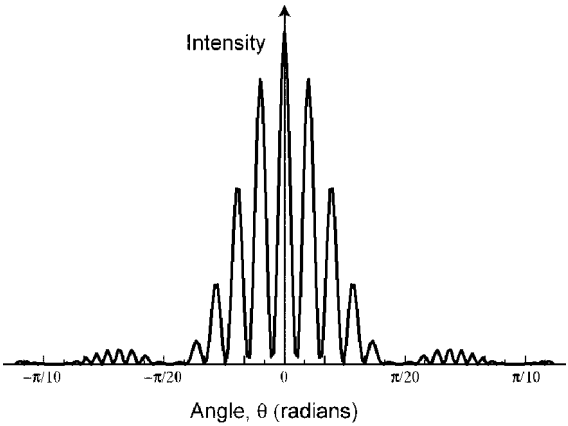


Figure 1.24 Young was the first to investigate the two slit interference patterns and a typical variation of intensity that would be seen as a function of the angle θ is shown for slit separation, $d = 1.5$ mm and slit width, $a = 6\mu m$.

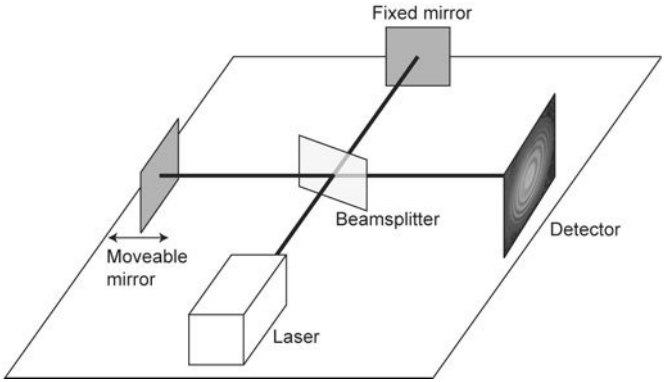


Figure 1.25 An important configuration in which two light beams are made to interfere is called the Michelson interferometer. Here, collimated laser light is partially reflected off the beamsplitter front surface and directed towards a moveable mirror while that part of the incident beam which is transmitted through the beamsplitter is directed towards a fixed mirror. After reflection from the two mirror surfaces, the beams are recombined at the beamsplitter and form an interference pattern at the detector.

1.3.5 Optical filters

It is often the case that certain light wavelengths have to be separated selectively from other wavelengths within an optical system. The most common example is in the theatre where white floodlights have gelatin sheets in front of the lamp to select different colours. These gelatin (or coloured glass) type filters have transmission characteristics, as shown in Fig. 1.26(a), which also allow through infrared wavelengths. Other types of filter are shown in Fig. 1.26(b) and a typical use of these filters is illustrated in Fig. 1.26(c). In fluorescence microscopy, for example, there are 'excitation filters' which allow through wavelengths smaller than the fluorescence wavelengths and 'barrier filters' which allow through only the longer fluorescence wavelengths.

However, especially in research work, there is often a need for narrow-band filters which transmit a narrow range of wavelengths and the interference effect described above may be used to create such 'interference filters'. If a semi-transparent metal film is evaporated onto a glass substrate, followed by a thin layer of dielectric material such as cryolite and a second metallic layer, these layers can then be protected by another thin glass plate. Assuming that the thickness of the dielectric film is d and its refractive index is n, those wavelengths which interfere constructively at normal incidence (and hence which result in most light being transmitted) will be given by

$$\lambda_{int} = \frac{2nd}{m} \qquad [1.36]$$

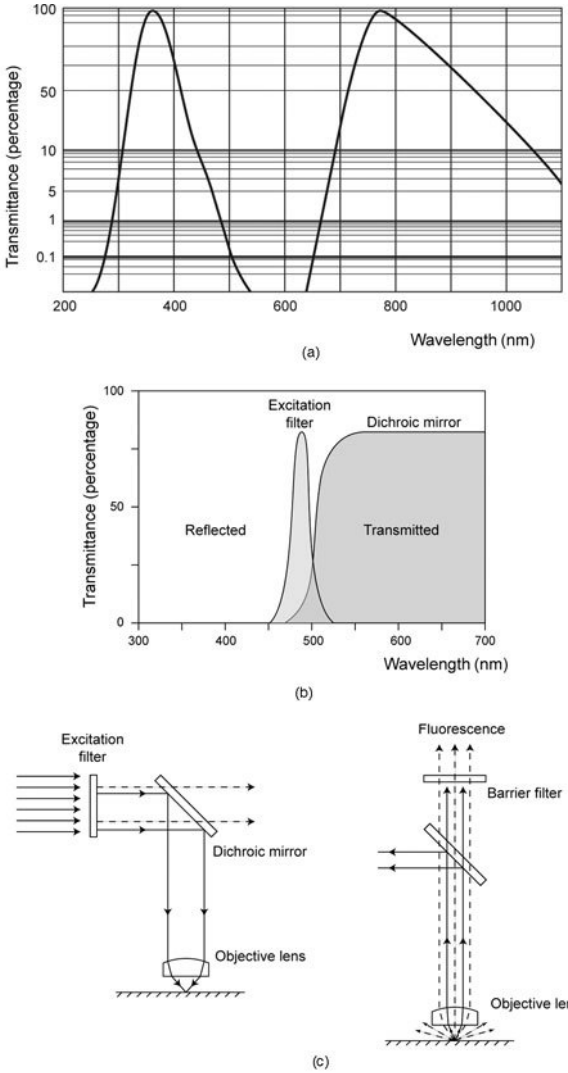


Figure 1.26 All materials act as some kind of filter for light transmission. No material can be perfectly transparent at all frequencies (because of the regions of absorption where there are large refractive index variations, see Fig. 1.12(a)). (a) A typical coloured glass transmission characteristic (BG3 from Spindler and Hoyer) shows not only transmission in the green but also in the infra-red. (b) A schematic diagram shows the narrow transmission characteristic response for excitation filters and also the long wavelength transmission characteristics of a dichroic mirror. (c) By careful design, these filters can separate out the light of different wavelengths. For example, the low wavelength light from the excitation filter is reflected from the surface of the dichroic mirror beamsplitter and directed onto the sample. If the sample fluoresces, the long wavelength fluorescence radiation passes through the dichroic mirror and, before detection, any further unwanted excitation radiation can be removed by an extra barrier filter.

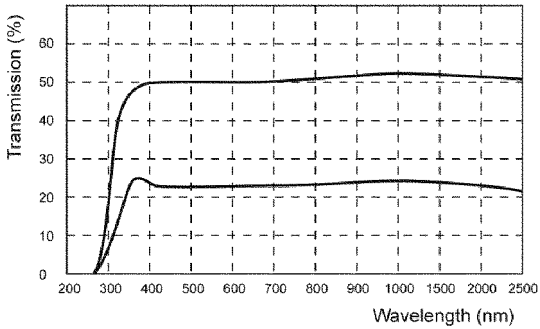


Figure 1.27 The neutral density filter has a uniform transmission over a wide wavelength range (visible and infrared) and, hence, these filters give selectable attenuation of incident laser light in many microscopes.

where m is the order of the interference ($m = 1, 2, 3, \ldots$). If the dielectric film thickness is sufficiently small, only a few wavelengths, λ_{int}, within the optical range will give maximum transmission. The unwanted optical wavelength can always be removed using a suitable coloured glass or gelatin filter. Nowadays, very narrow filter pass-band wavelength ranges (the narrowest being around 1.5 nanometres) centred on any wavelength within the optical range, can be fabricated reliably. Typical filter spectral responses (or transmission characteristics) are shown in Fig. 1.26(b) and an example of their use in a microscope application is shown in Fig. 1.26(c).

However, probably the most common type of filter used in microscopy is the neutral density filter. Neutral density filters may be fabricated by the vacuum-evaporation of metallic films onto glass substrates. The transmission characteristic of this type of filter is shown in Fig. 1.27 and it is used to uniformly reduce the intensity of incident light over the whole of the optical spectrum. Hence it may be used to reduce the laser intensities within an optical system by a calibrated amount. Often, different optical filters are placed on either a manual sliding mechanism or alternatively a multi-position filter wheel (which may be stepper motor controlled by a host computer).

1.3.6 Diffraction phenomenon

Diffraction is a phenomenon related very closely with interference. Classically it is described as being the result of interference between different parts of the light wavefront (rather than interference between two discrete light sources). Huygens successfully predicted diffraction effects by assuming that the light wavefront could be imagined as an infinite set of infinitesimally small secondary wavelets, as shown in Fig. 1.28(a) and (b). Even when a small slit is fabricated for illustrating the Young's interference experiment, the slit size will be many light wavelengths (hundreds of microns) in width. Therefore, if monochromatic

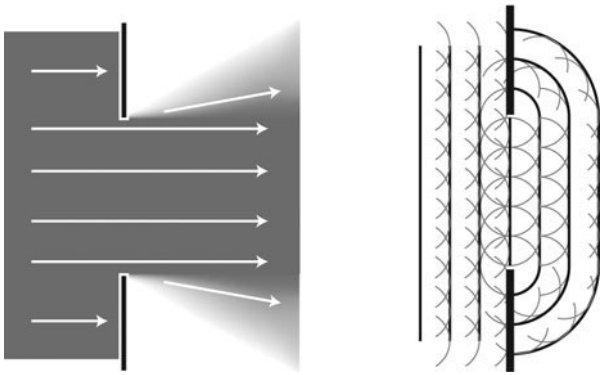


Figure 1.28 Whenever light passes through an aperture it shows the phenomenon of diffraction. Qualitatively, diffraction is the spreading out of the light into the geometrical shadow region. Huygens explained the effect in terms of secondary wavelets, which acted as sources of coherent light within the aperture (or slit in this case). The light from each of these secondary wavelets interferes to create the diffraction effect.

light is passed through only one slit, a series of light and dark bands will still be detected on a viewing screen, and the spatial variation in intensity across the resulting dark and light bands is shown in Fig. 1.29. This diffraction pattern is described by a mathematical expression involving the slit width, a, which differs only slightly from the equation for Young's slit interference, equation [1.35]:

$$I = I_0 \frac{\sin^2 \phi/2}{\phi/2} \text{ where } \phi = \frac{2\pi}{\lambda} a \sin \theta \qquad [1.37]$$

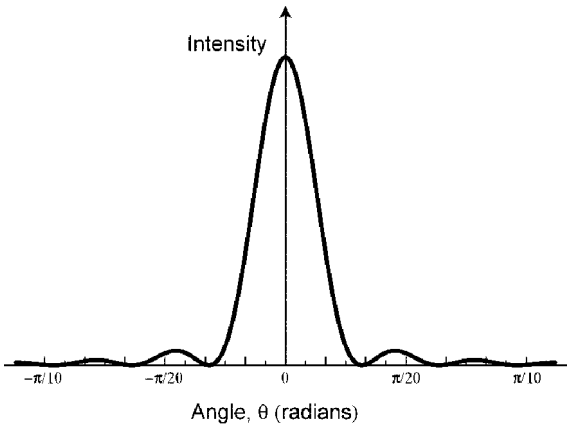


Figure 1.29 The variation in intensity seen on a screen as a function of the off-axis angle, θ, is shown for a slit width, $a = 9\mu m$. Note the similarity of the shape of the diffraction plot to the variation of the peak intensities of the interference fringes in Fig. 1.24.

The intensity, I, is at a point which is distant, d, from the optical axis on a screen placed a distance D from the slit and parallel to the slit. In the expression above, the angle θ is given by $\tan^{-1}(d/D)$.

Whenever light passes through some kind of aperture (or opening) in an optical system, the ultimate clarity, or spatial resolution, of the resulting image will be limited by this diffraction process. As optical lenses are circular, the most significant type of aperture is the circular aperture and the most important expression for microscopy (see Section 1.4.3) is the equation for the Airy light distribution

$$I = \left[\frac{2J_1(z)}{z}\right]^2 \qquad [1.38]$$

where $J_1(z)$ is the Bessel function[6] of order 1 and

$$z = \left(\frac{2\pi}{\lambda}\right) h' \sin U'_{max} \qquad [1.39]$$

where h' is the distance off-axis and U'_{max} is the semi-angle of the emergent light cone. The shape of the resulting light distribution pattern for the single slit looks remarkably similar to the intensity distribution through the centre of the circular image for the circular aperture. The image of a circular aperture is shown in Fig. 1.30. The bright central disk of the circular aperture diffraction pattern is called the Airy disk and, as the first minimum intensity occurs at $z = 1.22\pi$, the first zero occurs at a distance off-axis given by

$$h' = 0.61\frac{\lambda}{\sin U'_{max}} \qquad [1.40]$$

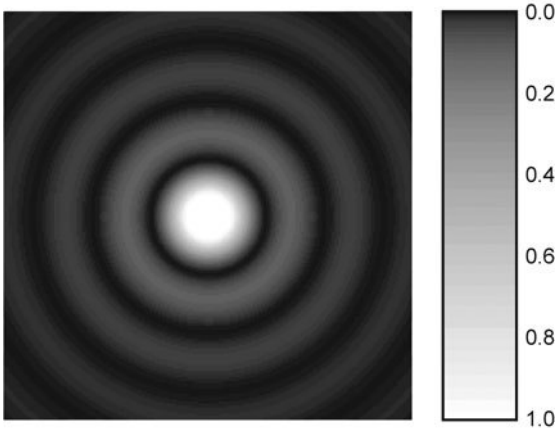


Figure 1.30 If light from a point source passes through a circular aperture, the variation in intensity seen on a screen will be circularly symmetric. A plot of the intensity through the centre of the image is very similar to the single slit, intensity variation plot, see Fig. 1.29. A bright central spot is surrounded by fainter circular fringes. This is called the Airy light distribution pattern.

1.3.7 Absorption and attenuation of light

Some materials exhibit transparency – light propagates through these materials with small losses in intensity. The properties that determine material transparency are absorption and attenuation. Human eyes are responsive to radiation from only a small part of the electromagnetic spectrum, i.e. a narrow range of wavelengths between 380 nm (blue) and 750 nm (red). Hence, when a material is apparently 'transparent', it is our evaluation of the propagation of these wavelengths through that material (whereas the same material may be opaque at other wavelengths). Similarly, a material, which is opaque at optical wavelengths, may be transparent in another wavelength range.

There are many different processes that could take place between light and a particular material but a simplification can be made which encapsulates all of these processes, as shown in Fig. 1.31. The assumption is that, when light passes through a small thickness of material, δl, the reduction in light intensity, δI, is proportional to the incident intensity and the material thickness. Mathematically,

$$\delta I = -\kappa I \delta l \qquad [1.41]$$

where the constant, κ is called the absorption coefficient for that material. By integrating over the path of the light wave, L, it can be shown that emerging intensity, I (which is a function of wavelength), is related to the incident intensity, I_0, by the Beer–Lambert Law expression

$$I(\lambda) = I_0 e^{-\kappa(\lambda)L} \qquad [1.42]$$

Sometimes the optical density (*O.D.*) is used to specify the absorption of the material of interest and the optical density is defined by

$$O.D. = -\log_{10}\left(\frac{I(\lambda)}{I_0}\right) \qquad [1.43]$$

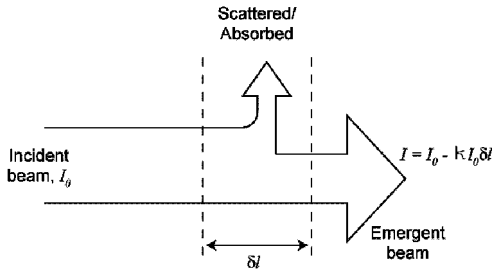


Figure 1.31 There are many processes which can lead to attenuation of the light beam as it travels through a material. These processes tend to depend on the thickness of the material, δl, and the original intensity of the light beam, I_0. If the emergent beam, I, can be written as $I = I_0 - \kappa I_0 \delta l$, where κ is an attenuation characteristic of that material, the absorption processes lead naturally to an exponentially decreasing light intensity with increasing material thickness.

It follows therefore that the connection between the absorption coefficient and the optical density for a material of thickness, L, is

$$O.D. = \frac{\kappa L}{\log_e(10)} = 0.434\kappa L \qquad [1.44]$$

Note that, although glass is perfectly transparent at optical wavelengths, it shows strong absorption bands outside the visible range, in the ultraviolet, as shown in Fig. 1.12(b), where the refractive indices increase significantly. Hence the absorption coefficient values must be referenced to the specific wavelength of interest. Note that the transmissivity, T, of the light flux through a parallel-sided, absorbing block of material of thickness, L, will therefore be given by

$$T = (1 - R_1) \cdot e^{-\kappa L} \cdot (1 - R_2) \qquad [1.45]$$

where R_1 is the reflectivity at the front face of the block and R_2 is the reflectivity at the rear face of the block.

Scattering is caused by variations of the refractive index within the material on a length scale smaller than the wavelength of the incident light, and this scattering process may be due to impurities, defects or inhomogeneities. Note that scattering affects the light beam in the same way as absorption. Hence it may be described by a similar equation. If there are N scattering centres per unit volume and σ_s is the scattering cross-section of each scattering centre, the intensity of the light beam also decreases exponentially as it propagates through the material.

$$I(z) = I_0 e^{-N\sigma_s z} \qquad [1.46]$$

George Stokes discovered that materials exhibiting the luminescence effect produced luminescence light with larger wavelengths than the incident light absorbed by the material. This unexpected effect was termed the 'Stokes Shift'. Even Maxwell realised that there was a fundamental problem with the classical electromagnetic theory of light's treatment of the interaction with materials, and it was only when the quantum theory was finally proposed that these issues were resolved.

The quantum theory took 25 years to unfold at the beginning of the twentieth century. In 1900, Planck hypothesised that, in order to explain the shape of the black-body radiation curve (shown in Fig. 1.32), the vibrating molecules in any heated material can only vibrate with certain discrete amounts of energy which he called quanta. He further stated that the minimum energy of vibration, $E_{min} = hf$, where f was a natural frequency of oscillation of the electric charges within the material and h was a fundamental constant, which is now called the Planck constant. In 1905, in order to explain the photoelectric effect, Einstein,[7] reasoned that light must also be emitted in discrete packets of energy, which he termed 'photons' and that the energy possessed by each photon must be

$$E = h\nu \qquad [1.47]$$

where ν was the frequency of the light. The photoelectric effect, shown in Fig. 1.33 and discussed in the next section, could not be explained by Maxwell's

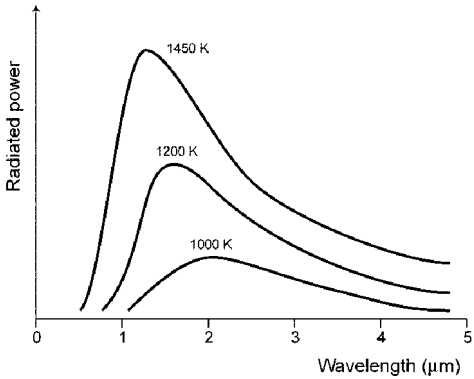


Figure 1.32 The power radiated from objects at a particular temperature will follow the form of the black-body radiation curves as shown. The higher the temperature, the greater the area under the curve and the peak emission moves towards lower wavelengths. The relationship between object temperature, T (in degrees Kelvin), and the wavelength of peak power emission, λ (in mm), is given by Wien's Law, $T = 2.898/\lambda$. Before Planck's quantum idea was accepted, the classical view of radiation predicted that the radiated power should continue to increase as the wavelength decreased – called the Ultraviolet Catastrophe.

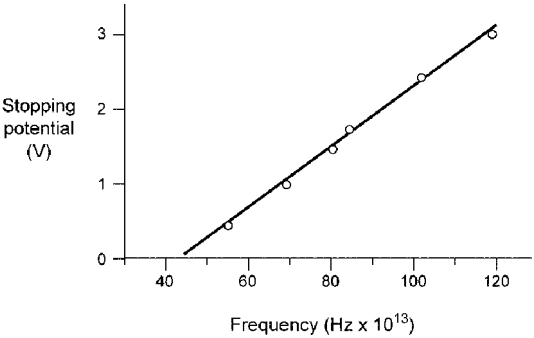


Figure 1.33 Einstein gave the first convincing argument for the photoelectric effect and the plot of the stopping potential required to prevent an electron being released from a material, which was irradiated with light of frequency, ν, could only be explained by a quantum approach.

classical EM wave theory but was a triumph for the photon description of light.

1.3.8 Quantum mechanics and Bohr Theory

Around 1900, the prevalent view of the structure of the atom was of a mini-solar system with a central, heavy, positively-charged nucleus and with lightweight, negatively-charged electrons orbiting the nucleus. However, as a result of

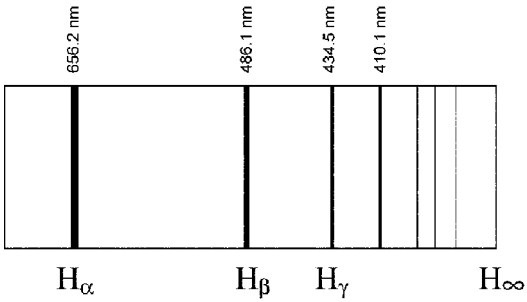


Figure 1.34 Another piece of crucial evidence for the quantum theory was the presence of discrete emission lines in the spectra from regions of excited hydrogen gas. Classical ideas could not explain these lines, but Bohr was able to predict their wavelengths using a model which allowed only certain electron orbits within the atom.

precise measurements of the wavelengths of the discrete emission lines in the radiation from excited gases (see Fig. 1.34), it became clear that the simplest model of these orbiting electrons had to be overhauled. Although a number of physicists had derived empirical formulae to predict the locations of the different series of emission lines in the hydrogen emission spectrum, these mathematical formulae did not explain why these lines should be present. In 1913, the Danish physicist, Niels Bohr, proposed two fundamental assumptions to explain the observed hydrogen emission spectrum.

1. The electrons in an atom can occupy only certain discrete orbits or quantised states. These states have different energies and the lowest energy state is called the ground state (which is the normal atomic state).
2. When an electron undergoes a transition from one state to another, it can do so by emitting or absorbing radiation. The frequency, ν, of this radiation is given by

$$\nu = \Delta E/h \qquad [1.48]$$

where h is Planck's constant (6.63×10^{-34} J.s) and ΔE is the energy difference between the two states involved.

These assumptions implied that, in terms of emission and absorption processes, light had best be thought of as a stream of discrete bundles of energy called 'photons'. By analysing the light from a hydrogen discharge tube after dispersion by a diffraction grating, the characteristic line wavelengths allow the quantum states of hydrogen to be inferred, using Bohr's postulates. These allowable orbits or states may be likened to standing wave patterns on a string. De Broglie, in his doctoral thesis, proposed that every electron has wavelike properties and the wave has a wavelength given by $\lambda = h/p$ where p is the momentum of the electron. The only way that an electron could maintain a particular orbit is when the circumference of the orbit equals an integral number

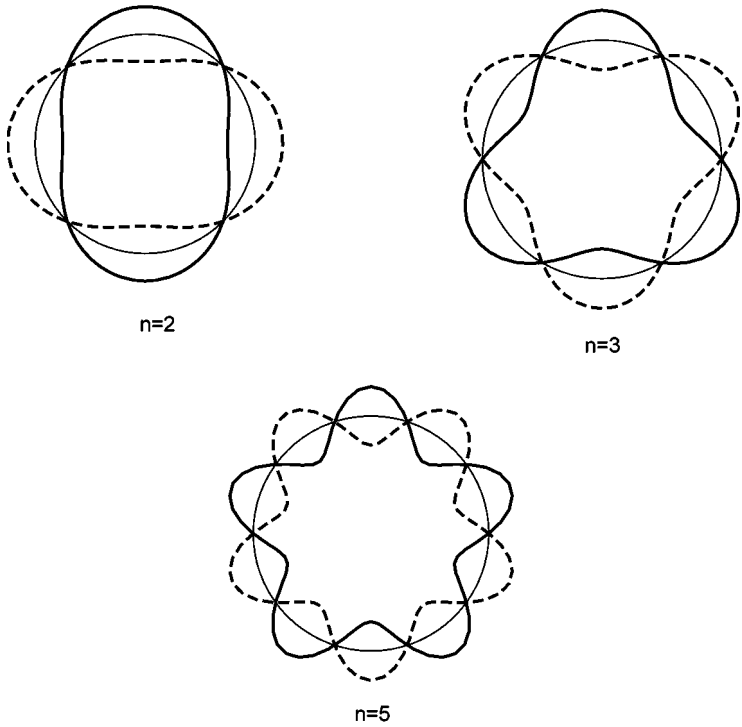


Figure 1.35 The only electron orbits allowed in Bohr's theory of atomic structure were those whose circumferences were a discrete number of electron wavelengths, λ_e. In this way, considering the electron as a wave-like entity, standing waves would be produced for radii, r, where $2\pi r = n\lambda_e$ as shown. For other orbital radii, the electron waves would destructively interfere and no standing waves could be created. Higher discrete n values corresponded to states of higher electron energy.

of wavelengths, as shown in Fig. 1.35 (otherwise the electron waves would destructively interfere).

Because of this discrete energy state picture of the electrons in an atom, there are several interesting phenomena that can occur (besides absorption) when a photon is incident on an atom. Consider the photon to have a frequency, ν, so that its quantum of energy is given by $h\nu$. Referring to Fig. 1.36(a), if this photon energy is insufficient to raise an electron into a higher allowed energy state, the atom stays in its original state and the photon is 'elastically scattered'. When the wavelength of the photon ($\lambda = c/\nu$) is large compared to the size of the atom, this elastic scattering can be described by classical electromagnetic theory and corresponds to Rayleigh Scattering.[8] The probability of light of wavelength, λ, being scattered in this way is proportional to $1/\lambda^4$.

If the energy required to raise the ground state electron into the first excited state is ΔE and the incoming photon has exactly this energy, 'resonance

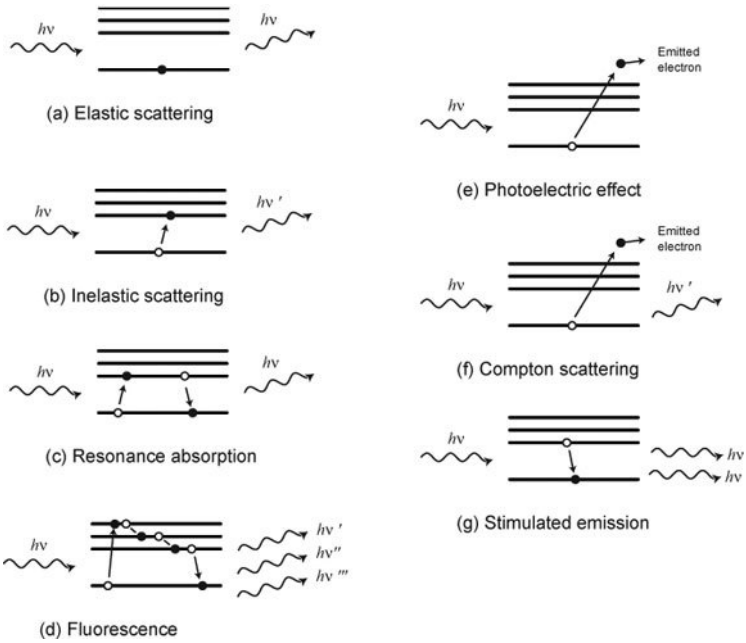


Figure 1.36 The allowed electron energy states are represented by lines in an energy state diagram, and lines which are higher in the diagram represent higher energy states. The lowest line is the ground state energy and electrons must be given energy if they are to move into higher energy levels. The different types of light–material interactions, described in the text, are illustrated with these energy level diagrams. (a) elastic scattering, (b) inelastic scattering, (c) resonance absorption, (d) fluorescence radiation, (e) ionisation and the photoelectric effect, (f) Compton scattering, and (g) stimulated emission.

absorption' takes place where the electron hops into the excited state but then falls back into the ground state. When the electron falls back to the ground state, the electron emits a photon with energy ΔE just like the incoming photon (but with a phase uncorrelated to that of the incoming photon) – this is called 'spontaneous emission'.

However, if the energy of the photon is sufficiently high to raise the electron into a higher energy state, as shown in Fig. 1.36(b), the photon is said to be 'inelastically scattered' and the photon will lose an amount of energy, ΔE, to the atom. The emerging photon from the interaction will therefore have a different frequency, ν_f, given by

$$h\nu_f = h\nu - \Delta E \qquad [1.49]$$

and this effect was first described by C.V. Raman and is referred to as Raman Scattering. Raman microscopy has an important place in modern microscopy and is discussed further in Chapter 5.

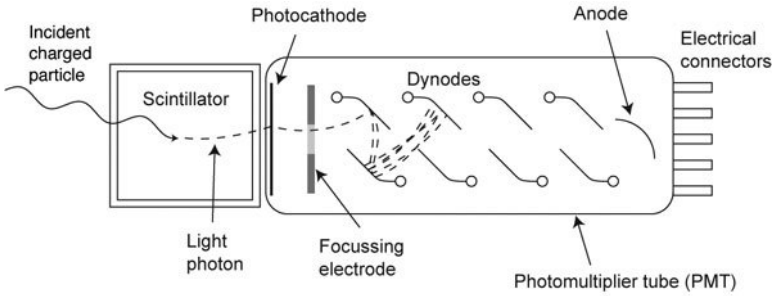


Figure 1.37 The photomultiplier tube (PMT) has been one of the most important low light level detectors in recent times. A thin film photocathode is deposited on the inside of a glass front surface, which emits an electron when a few photons are incident. A series of metal grids, called dynodes, are connected to ever-increasing positive potentials and the highest potential is on the anode at the end of the tube. The electron(s) emitted from the photocathode are accelerated to the nearest dynode where they knock out more electrons. An avalanche effect takes place as ever-increasing numbers of electrons travel along the dynode chain. A brief duration pulse is therefore generated at the anode, whose size indicates the number of photons originally incident on the photocathode. The PMT can also be used to detect high energy charged particles, provided that a scintillator, or other device, is attached, which produces photons when the particle traverses the scintillator.

If the incoming photon has even greater energy, it can raise the ground state electron into higher energy level excited states, as shown in Fig. 1.36(c). Within a very short time (typically 10^{-8} seconds), the atom relaxes back into the ground state via a number of intermediate excited states and, as the electron hops into lower states, photons of lower energy are generated. This process is called 'fluorescence'. In some materials, the atoms can exist in the excited states for milliseconds, seconds or even longer and in this case the process is called 'phosphorescence' (as in many oscilloscope displays).

With incoming photons of even higher energy, the photon will eventually release the ground state electron from the atom (with no photon emitted). The incoming photon has 'ionised' the atom. This mechanism also explains the 'photoelectric effect' in which an electron is ejected from a material when ultraviolet light is incident on the material. The photoelectric effect is described by the equation

$$E_e = h\nu - W \qquad [1.50]$$

where W is called the 'work function' of the material and it represents the minimum energy required to release an electron from that material. By conservation of energy, if the photon is more energetic than W, the emitted electron will have a residual energy, E_e. This effect underpins the operation of one of the most sensitive optical measurement devices – the photomultiplier tube (PMT) – shown schematically in Fig. 1.37.

Incident photons strike the photocathode material deposited as a thin film over the glass window of the PMT and any photoelectrons liberated from the photocathode are attracted to and accelerated by the first of many dynodes (because of the positive potential on these dynode terminals). On striking a dynode, more electrons are liberated for each electron absorbed. Hence, an avalanche effect, of ever-increasing numbers of electrons, propagates along the dynodes until reaching the final anode. The accumulation of charge at the anode creates a fast voltage pulse, which can be further processed by appropriate electronics to determine the original number of photons at the photocathode (determined by the size of the voltage pulse). Nowadays, individual incident photons can be monitored by the most sensitive PMTs and some microscopy systems have a 'photon counting' mode of operation. Note that if the PMT is used as a detector, it is a point detector (giving a single measure of the number of photons striking its surface within a short time period). Hence a 2D image field must be created by raster scanning the PMT over the sample in some way, as described in more detail in Chapter 3.

Although the discussion above has been solely in terms of incident energetic photons interacting with atoms to produce these various effects, the atoms can also be excited by collisions with particles or other atoms. Hence, these processes can be initiated without incident photons being involved at all.

1.3.9 Sources of EM radiation

Objects in outer space either generate electromagnetic radiation (e.g. stars) or are observable because they reflect electromagnetic radiation from nearby stellar sources (e.g. satellites, the Moon, gaseous nebulae). In microscopy, few objects are self-luminous (although many polymers exhibit auto-fluorescence). Therefore, most specimens must be illuminated by powerful light sources so that the light reflected from their surfaces, or the light transmitted through them, can create images with good contrast (and high signal-to-noise ratios). Another necessary condition is that the specimen be uniformly illuminated to achieve the best results.

In the earliest days of microscopy, apart from using directed solar radiation, candles were the only means of illumination of specimens – hence the original illumination levels being measured in units of 'candle power'. The creation of reliable, intense light sources for microscopy was influenced by Edison's development of incandescent light bulbs (see Fig. 1.38). The illumination of the specimen may be enhanced by locating the source at the focus of a spherical, metallic reflecting surface.

The most common source for the latest microscope systems is the laser. The laser is a device based upon the principle of stimulated emission, which was discussed in the last section. Its name is derived from *l*ight *a*mplification by the *s*timulated *e*mission of *r*adiation. If an atom is already in an excited state and it interacts with a photon whose energy is the same as the energy difference

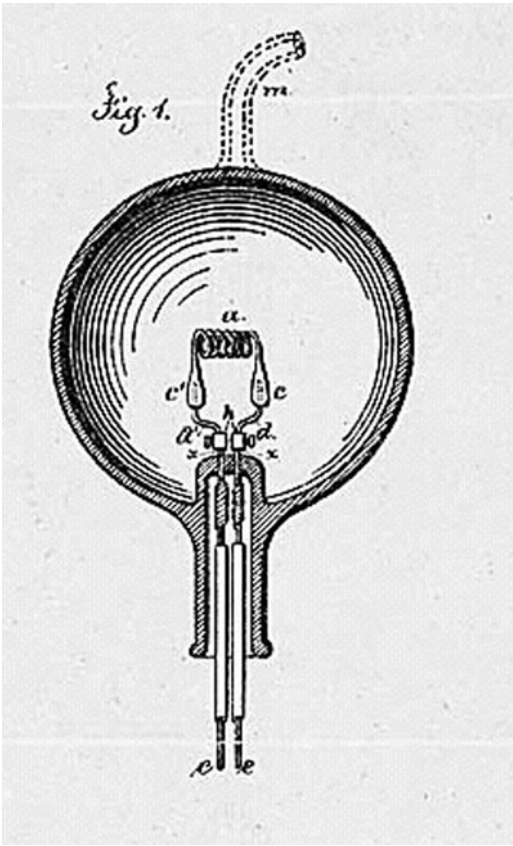


Figure 1.38 The first reliable source of artificial light was Edison's 'Electric Lamp' shown in this original patent 223,898 from 27 January 1880.

between the excited state and the ground state, the process of 'stimulated emission' can occur. The photon emitted, as the electron falls back to its ground state, is in phase with the incident photon. If this process could be encouraged in billions of atoms, gradually a high energy, coherent light beam would be developed. Therefore the basic technique is for the atoms within the material inside the laser to be put into an excited state by a high energy optical flash before the incident light comes along to generate the stimulated emission process (see Fig. 1.39). The electron in the high energy state de-excites emitting a photon with the same energy as the incoming photon and in phase with the incoming photon. These photons bounce back and forth within an optical cavity and eventually leak out of one end of the cavity through a half-silvered mirror surface. Hence the incoming radiation is amplified and the light produces an output beam that is both coherent and collimated.

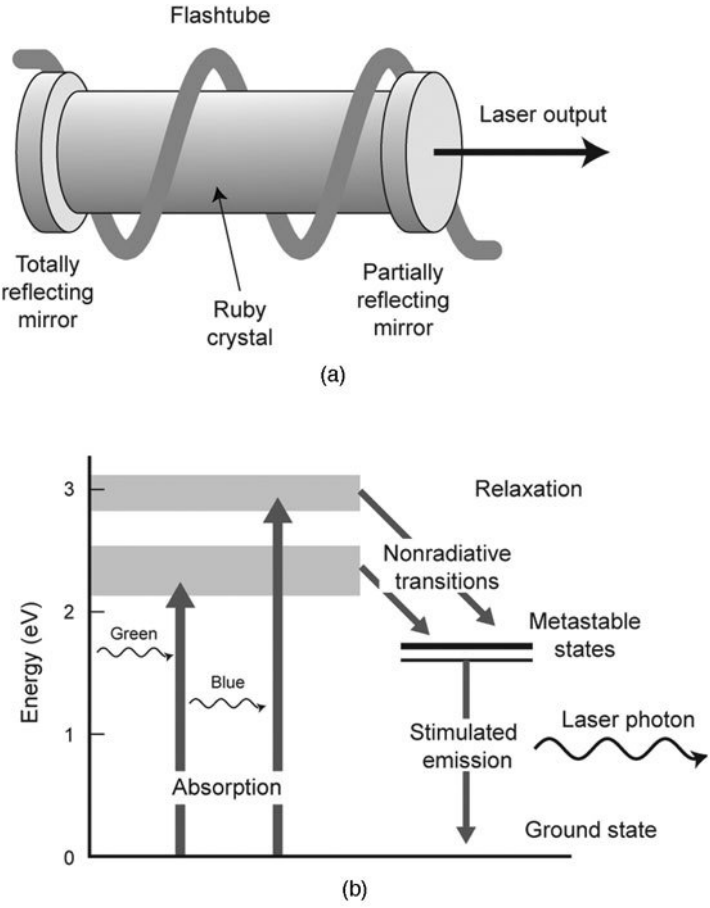


Figure 1.39 A schematic diagram of the ruby laser operation where an intense optical flash raises the electrons in the ruby into higher metastable states prior to light entering the ruby crystal. As this light travels backwards and forwards within the ruby crystal, the electrons de-excite by stimulated emission and high-energy coherent radiation is created within the ruby, which gradually leaks out of the partially reflecting mirror at the end of the ruby crystal.

1.4 Elements of microscope designs

So far in this chapter, EM waves have been discussed in relation to their interaction with planar and non-planar interfaces and apertures. Microscopes could be designed to create magnified images of small objects using mainly reflecting optical structures (see Section 1.4.5). However, this type of microscope tends to be used nowadays only for specialist ultraviolet or infrared applications (because of the high attenuation of ultraviolet and infrared EM

waves by standard glass lenses). Most microscopy systems consist of combinations of refracting lenses and these are discussed in more detail below.

1.4.1 The magnifying lens – or 'simple microscope'

Before discussing man-made lenses, it is worthwhile considering the natural lens that occurs in the eye. A cross-section of the eye is shown in Fig. 1.40. The muscles surrounding the lens are able to alter the lens curvature and hence the focussing ability of the lens. This 'accommodation' of the eye enables us to focus on objects at different distances. The spectral sensitivity of the eye naturally determines what we regard as the optical part of the electromagnetic spectrum, i.e. from 380 nm to 760 nm. By considering the ray paths, it can be seen that an inverted image of the external scene will be formed on the light sensitive cones of the retina. (Incredibly, our brain processes this inverted image to create a non-inverted picture of the world in our consciousness!)

The smallest object detectable by the unaided eye must be placed as close to the eye as possible. But, depending on the age of the viewer, it is not possible to see clearly an object if it is closer than the 'least distance of distinct vision' or

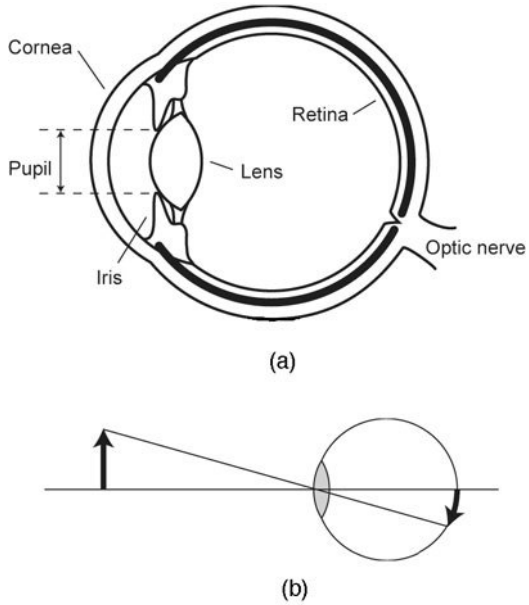


Figure 1.40 (a) The cross-section of an eyeball shows the lens supported by the ciliary muscles and the lens curvature is controlled by these muscles. The lens images the scene onto the light sensitive rods and cones on the retina at the back of the eyeball. The pupil regulates the amount of light entering the eye. (b) Note that an inverted image is created on the retina.

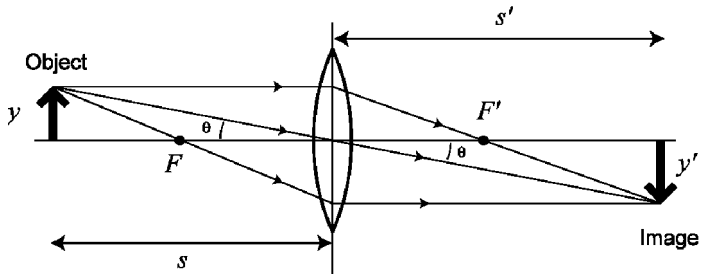


Figure 1.41 Using the same ray-tracing principles as in Fig. 1.22, the image formed by an object beyond the focal point of a thin convex lens is shown. The thin lens formula, $1/s + 1/s' = 1/f$, where f is the focal point distance, can be used rather than these ray-tracing diagrams to deduce the image location (and its size using the equation for magnification).

near point of the eye. The near point is usually assumed to be 25 cm from the eye. Given the size of the retinal cones at the back of the eye, an object approximately 300μm in height can be distinguished comfortably at this distance. Hence the need for a magnifying lens or microscope system to enable the detection of smaller objects.

It is well known that the Greeks understood the properties of mirrors and glasses which could condense solar radiation into a burning beam. The Vikings may even have had lens knowledge (on the basis of evidence that glass artefacts, shaped like a lens, have been found in excavations) but there are no written texts from that period to verify this. However the first person to study lenses and the action of lenses was the Islamic scholar/philosopher, Alhacen in the eleventh century.[9]

Ray tracing can be used to describe the formation of images by a magnifying lens, as shown in Fig. 1.41. However, it is more convenient and quantitative to work out the image position by using the lens maker's formula. For a thin glass lens in air, the formula is given by

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} = P = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad [1.51]$$

where P is the power of the lens (in diopters or m^{-1}), f is the focal length, s and s' are the object and image distances along the lens axis respectively, n is the refractive index of the glass and r_1 and r_2 are the radii of curvature of the front and back lens surfaces. In order to apply this formula correctly for both convex and concave lenses, an appropriate sign convention must be followed and the convention is shown in Table 1.3.

Images in optical systems may be virtual or real. A virtual image is an image which appears to be located at a point in space, but, if one were to place a screen at that place, no image would be seen on the screen. For example, when we look into a mirror, the image of our face appears to be located an equal distance behind the mirror. It is called a 'virtual' image, however, because it would not

appear on a screen placed behind the mirror. In contrast, a real image would be seen on a screen placed where the image appears to be located. For instance, in every camera, a real image is formed on either the recording film or, in a modern digital camera, a sensitive large area sensor like the charge coupled device (CCD) discussed in Chapter 2.

Images may be conveniently located in a ray diagram by using any two principal rays. The point from which these rays diverge or appear to diverge is the image point. For a lens, there are three principal rays to take into consideration: the parallel ray (parallel to the axis); the focal ray, through the second focal point; and the central ray, through the centre of the lens. A positive or converging lens is a lens that is thicker at the middle than at the edges and a negative or diverging lens is one which is thicker at the edges than the middle, as shown in Fig. 1.42.

The magnifying lens consists of a convergent lens (or a convergent system of lenses) of short focal length. (This type of lens is actually the ocular or eyepiece lens of a compound microscope.) The object to be examined is placed within the focal distance so that a magnified, upright, virtual image is formed approximately at a distance at which the eye can see it most distinctly. The 'magnifying power' or 'lateral magnification' is a measure of the effectiveness of a magnifying lens. It is defined as the ratio of the apparent size of the image to the actual size of the object, when placed at the least distance of distinct vision.

The lateral magnification of any thin lens is therefore given by

$$M = \frac{y'}{y} = -\frac{s}{s'} \qquad [1.52]$$

where y' is the image size and y is the object size. A negative magnification means that the image is inverted. However, in order to apply this relationship effectively, the correct sign convention must be applied depending upon the lens being used (see Table 1.3).

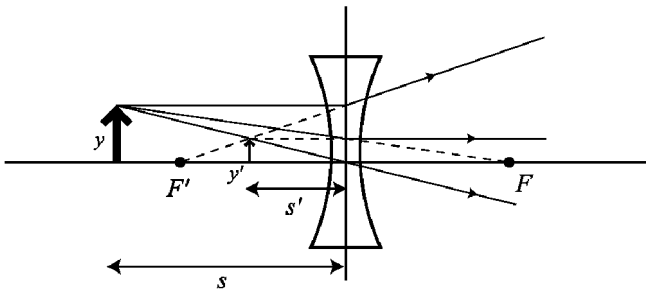


Figure 1.42 The ray-tracing principle can also be used for a thin concave lens to deduce the position of the de-magnified image. If the thin lens formula is used, care must be taken with the lens convention necessary to compute the image size and distance.

Table 1.3 Sign conventions for thin lenses and spherical interfaces

Parameter	Sign Convention +	–
r_1	Convex (front surface)	Concave (front surface)
r_2	Concave (back surface)	Convex (back surface)
f	Converging lens	Diverging lens
y	Erect object	Inverted object
y'	Erect image	Inverted image
M	Erect image	Inverted image

The variation of the magnifying power of a convex magnifying lens is shown in Fig. 1.43 as a function of the object distance, s. For comfort, the object distance is chosen so that the image distance of a magnifying lens is equal to the nearest distance of distinct vision. Hence, a useful expression for the magnifying power of the lens is given by

$$M = 1 + \frac{l_{near}}{f} \qquad [1.53]$$

where l_{near} is the nearest distance of distinct vision and f is the focal length of the lens.

Note that, although geometrical optics illustrates graphically how a lens can form a magnified image, it cannot say anything about the quality of the image that is formed. A basic knowledge of Maxwell's wave theory is needed to relate the characteristics of the lenses in an optical system to the spatial resolution and brightness contrast in the final image.

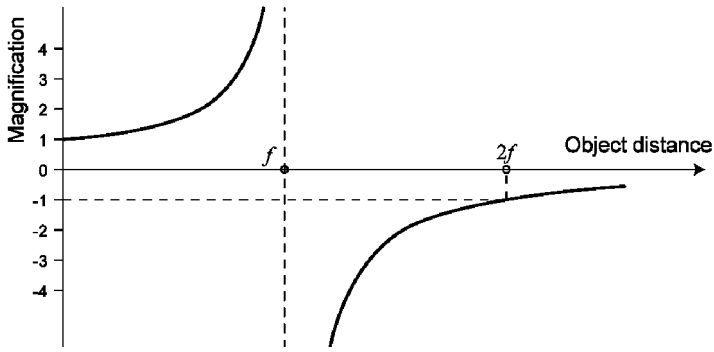


Figure 1.43 The thin lens formula (eqn [1.51]) and the expression for the magnification (eqn [1.52]) can be combined to predict the variation in the magnification with object distance for a convex lens. Note that, as the object approaches the focal point, the magnification rises sharply and, indeed, at the focus, no image will be formed because the light rays do not converge. The negative magnification for object distances greater than the focal length denote that the image is inverted.

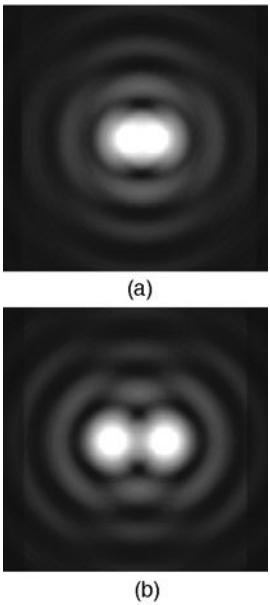


Figure 1.44 The first attempt to define the spatial resolution of an optical system was Rayleigh's criterion, which considered the overlapping of the Airy diffraction patterns produced by two, close, point sources. (a) When the peak intensity of the image of one source coincided with the first dark ring around the other source's image diffraction pattern, Rayleigh regarded these as two just resolvable sources. (b) If they were further apart, then one would expect their Airy patterns to be clearly distinguishable. However, now that images can be digitised, it is possible to distinguish two sources even when they are closer together than the Rayleigh criterion separation.

The most important physical effect determining the spatial resolution of images is the diffraction of light. As discussed earlier, when the light shining directly through a small, finite, point source (say a pinhole) is projected onto a screen, not only the high intensity central spot of light is seen but also a series of concentric light and dark rings, as shown in Fig. 1.30. If two pinholes were placed close together, as in Fig. 1.44, their central spots and diffraction rings would overlap with each other. The diffraction rings must clearly limit the spatial resolution of neighbouring objects. Hence, even when lenses are used to create an image, this diffraction phenomenon must also affect the image resolution. Lord Rayleigh proposed the connection between the wavelength of light, λ, the diameter, D, of the lens collecting light from the object and the minimum object separation, d_{min}, for two pinholes to be resolvable (i.e. distinguishable). He proposed that, when the bright Airy disk of one pinhole image overlapped the first dark ring of the neighbouring Airy pattern, the two pinholes would be just distinguishable. This rather crude measure implies that the limit of resolution is given by

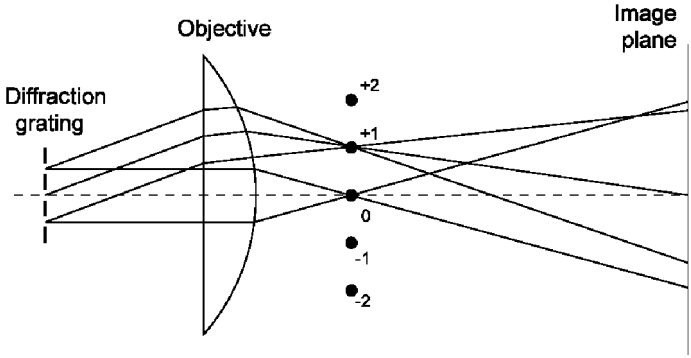


Figure 1.45 To illustrate that a lens in any optical system cannot give a perfect reproduction of the object, consider the imaging of a diffraction grating. The rays shown represent the zeroth and first order diffraction peaks. The objective lens focuses the parallel rays of each order onto the focal plane, but because of its finite size, it can only focus a few of the orders. Therefore the image is reproduced from only a few orders and can never be an exact facsimile of the object. Also note that orders can be filtered at the focal plane, as described later.

$$d_{min} = \frac{1.22\lambda}{2n \sin \alpha} \qquad [1.54]$$

where $n \sin \alpha$ is called the numerical aperture (NA) of the lens (see also page 37).

In order to reconstruct the original object, all of the light diffracted by the object would have to be recombined in the image. However, in practice, the objective lens of a microscope can only retrieve a fraction of the light waves (due to the finite diameter of the lens and its distance from the object, see Fig. 1.45). Another condition for faithful reproduction is that all of the light waves that are recombined should have traversed the same overall optical path length (so that the different components of light retain their same relative phase relationships). As neither of these conditions can be satisfied, no image is perfect and this Rayleigh limit exists for all optical systems.

1.4.2 The compound microscope

Roger Bacon, a Franciscan friar and teacher at Oxford University in the thirteenth century, is usually credited with the invention of spectacles (despite the fact that eye glasses were manufactured in Pisa, Italy in the late thirteenth century). Curiously, there is no historical evidence yet that anyone around this time thought of putting one lens in front of another in order to invent either the telescope or the microscope.

What is the difference between the lens combinations for a refracting telescope and a compound microscope? In a refracting telescope, the real image formed by the object lens (which has a long focal length, f_o) is formed approximately at the rear focal point. This corresponds to the front focal point of

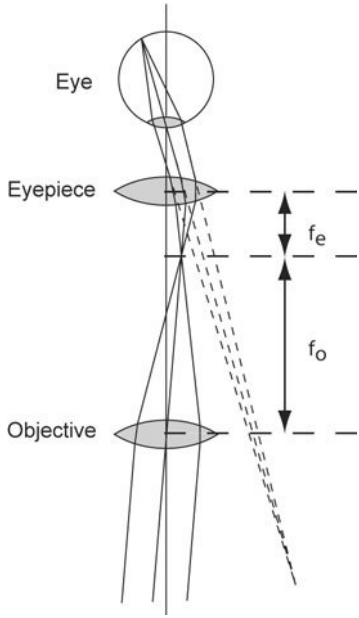


Figure 1.46 With a combination of two convex lenses, a refracting telescope could be built. Parallel rays from distant objects converge at the focal point behind the objective lens, which corresponds to the front focal point of the eyepiece lens. A long focal length, f_o objective is combined with a short focal length, f_e eyepiece and the magnifying power is given by $M = \theta_e/\theta_0 = -f_o f_e$. In effect the eyepiece acts like a simple magnifier for viewing the real image from the objective lens.

the ocular or eyepiece, as shown in Fig. 1.46. However, in a microscope, the object lens (objective) has a very short focal length. The object is placed so close to the front focal point of this lens that the image distance is much greater than the focal length. The real intermediate image is a magnified version of the object and it is at this intermediate field position that measuring grids are placed. The intermediate image is viewed through a magnifying glass called the ocular or eyepiece, which magnifies the image further, as in Fig. 1.47.

One might be tempted to magnify an object indefinitely by adding more lens stages but the French scientist, Abbé showed that, due to the wave character of light, it is only possible to distinguish details of a size down to the wavelength of the light used to illuminate the object. As discussed on page 58, the addition of more lens stages would lead to useless or empty magnification.

Whereas telescopes are physically large because they need to optimise the light collection efficiency from distant, low illumination objects, the microscope needs to optimise the spatial resolution. In all optical microscopes, the most intense, feasible light beam illuminates the object of interest (by transmission or reflection) and, by use of a physically small but large NA objective lens, good

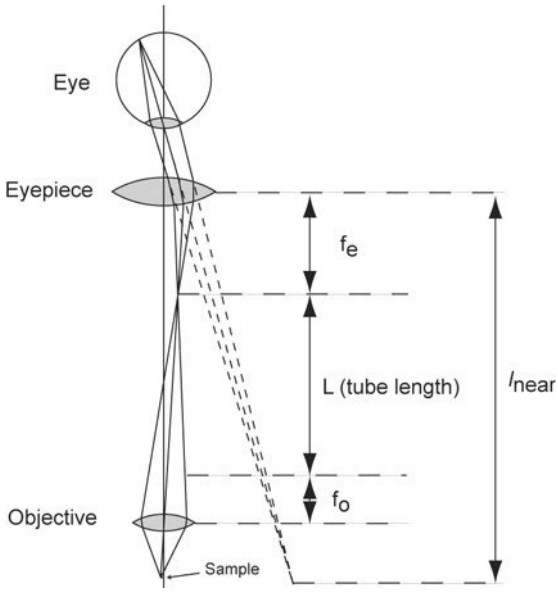


Figure 1.47 With a combination of two convex lenses, a compound microscope could be constructed. A very short focal length, f_o objective lens creates an image, which is some way from the focal point of the lens – a distance L from the focal point. The eyepiece lens of focal length, f_e views the real image created by the objective and the final image is placed at the least distance of distinct vision, l_{near}. The magnifying power $M = -Ll_{near}/f_o f_e$.

light collection efficiency is achieved. Even the man credited with the discovery of the microscope – a Dutch draper called Antoni van Leeuwenhoek – who lived in the seventeenth century, only devised a version of the simple microscope which consisted of a tiny bead lens held up to the eye in a metal plate. Such was the skill and patience of van Leeuwenhoek that he achieved significantly higher resolution with his simple microscope design than was possible with typical (more complex) compound microscope designs built around the eighteenth century! Galileo is usually credited with the discovery of the compound telescope around 1610.

However, the rather crude compound microscope designs which appeared in the seventeenth century did capture the imagination of the life scientists of the time. In 1660, Marcello Malpighi used his compound microscope to confirm William Harvey's theory of blood circulation by observing blood flow in capillaries. Following these observations, nearly every microscope sold for the next 200 years had, as an accessory, the fish- or frog-plate to observe blood circulation. The other subjects that occupied seventeenth century microscopists are all described in Robert Hooke's famous treatise *Micrographia* published in 1651.[10] The outstanding technical problems with the compound microscope were that the lenses suffered from two significant aberrations: chromatic

aberration and spherical aberration (see next section) and also the earliest decorative designs of microscope stand and barrel did not give good mechanical stability.

In a compound microscope, an objective lens is focussed onto the specimen, and an eyepiece views the image produced by the objective and creates a magnified real image, as shown in Fig. 1.47. Whereas the brightness of an image varies directly as the square of the numerical aperture (or as NA^4 in reflected light fluorescence microscopy), the brightness varies inversely as the square of the magnification, i.e. as M^{-2}.

The magnifying power of a compound microscope is the product of the lateral magnification of the objective lens and the angular magnification of the eyepiece lens:

$$M = -\frac{L.l_{near}}{f_o f_e} \qquad [1.55]$$

where L is the tube length (the distance between the second focal point of the objective and the first focal point of the eyepiece), l_{near} is the near-point distance and f_o and f_e are the focal lengths of the objective and eyepiece respectively.

When the coupling medium between a lens and the specimen is air, the maximum value of $n \sin \alpha$ is unity but in practice the actual NAs of these objectives rarely exceed 0.85. However, oil immersion objectives are often used in microscopy and, as shown in Fig. 1.48, the refractive index of the oil (and cover slip) produces an effective increase in the semi-angle of the cone of rays from the point source and hence a higher effective NA. The highest NA oil immersion objectives have NAs between 1.3 and 1.4 and the type of oil which gives optimum performance will also be defined. Note that objectives are designed with a particular coupling medium in mind (so for best performance they should not be used with a different coupling medium). Whereas objectives are designed to work with cover slips in transmission microscopy, metallurgical microscope objectives are designed to work without cover slips and hence the microscopist must be careful to choose the appropriate objective lens for the application.

The light collecting power of the objective lens (which is proportional to the NA of the objective) will determine the brightness of the image. When a lens is sharply focussed on a specified object plane, features lying above and below this optimum object plane also appear to be in acceptable focus. The range of distances over which features appear in acceptable focus is called the depth of field – but sometimes it is referred to, incorrectly, as the depth of focus. (The 'depth of focus' is, strictly speaking, the distance through which the image plane may be moved before the image becomes indistinct.) Hence the object plane can be considered a thin layer in which all points appear to be in focus. The depth of field, h, can be shown to be

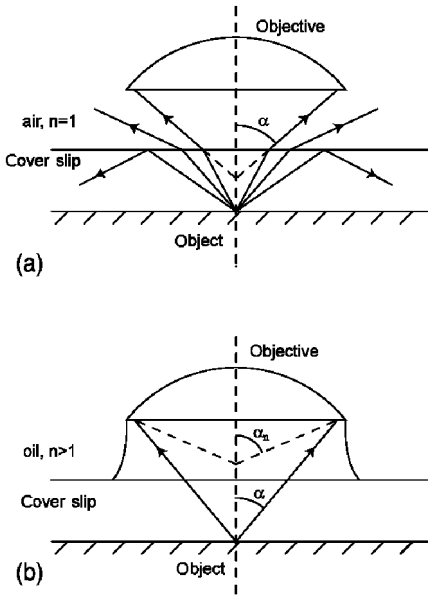


Figure 1.48 In some applications, cover slips are used to protect the specimen from damage. (a) Because of refraction, the range of angles at the specimen, over which radiation can be detected by the objective, is less than the acceptance cone of half angle, α, when the cover slip is in air. (b) However, when oil immersion objectives are used with the cover slip and the refractive index of the oil is matched to the refractive index of the cover slip, the acceptance cone is far greater.

$$h = \pm \frac{\lambda}{4n \sin^2\left(\frac{\alpha}{2}\right)} \qquad [1.56]$$

where n is the refractive index of the object space. Values of the depth of field (for an illumination wavelength of 550 nm) for different lens types are given in Table 1.4.

The effective field of view of a microscope depends on both the objective and the eyepiece lenses. The intermediate image formed by the objective is located around 18 mm below the rim of the microscope tube. The standard internal diameter of a microscope tube (23.2 mm) will ultimately restrict the size of this image but, in practice, it is the design of the eyepiece lens which limits the size of the intermediate image to around 18 mm diameter. Hence, if the magnification of the eyepiece is $\times 10$, the diameter of the field of view will be 1.8 mm. If the magnification is $\times 40$, the diameter of the field of view will be 0.45 mm. Microscopes with special wide-field tubes of 30 mm diameter are also manufactured and, by offering flat-field objectives and wide-field eyepiece lenses, such a system will be especially optimised for photomicrography.

Table 1.4 Depth of field values for different objectives

Numerical aperture (NA)	0.10	0.25	0.50	0.75	1.00	1.25
Depth of field (dry objectives) Air ($n = 1$) Distances in μm	56	8.6	2.0	0.81	–	–
Depth of field (oil immersion objectives) Oil ($n = 1.5$) Distances in μm	82	13	3.2	1.4	0.73	0.42

One crucial question in microscopy is this: what magnification is needed so that the microscopist can perceive all detail in the object? Ultimately, this depends upon the eye's ability to resolve detail. The unaided eye is able to distinguish two points if they subtend an angle of 1 minute of arc at the eye, which corresponds to a separation of 100 μm at the near point of the eye. But this is pushing things to the limit – it is better to have a separation of two or three times this distance to avoid eyestrain. The minimum magnification needed for the eye to just distinguish two points is therefore the ratio of this separation (say 300 μm) to the resolution of the objective lens. In reality, this means that the optimum magnification (or useful magnification) of the objective is between 500$\times$ and 1000$\times$. If higher magnifications are used, the image size will be increased without increasing the amount of detail and the quality is degraded because the inherent fuzziness of the Airy pattern from every point source within the image field will be seen by the microscope user. This excess or useless magnification is called empty magnification.

1.4.3 Aberrations and real lenses in compound microscopes

So far, ideal lenses have been discussed. Real lenses in real microscopes suffer from various aberrations or effects, which reduce the overall image quality. Care must also be taken to use the appropriate objective lens and eyepiece lens for the application. Hence, in this section, a number of these issues will be elaborated.

The introduction of brass for the microscope stand and barrel (or tube) in the mid-eighteenth century solved the stability problem. Unfortunately, the first lenses suffered from longitudinal chromatic aberration (i.e. different colours were focussed to different positions on the lens optic axis), as shown in Fig. 1.49(a), due to the variation of the refractive index, n, with the wavelength of light, λ. However, a clever way to correct for chromatic aberration (in telescope lenses) was eventually solved by the Dollond family. They used a combination of crown and flint glass to create an achromatic doublet lens, shown in Fig. 1.49(b), which is still the basis for this correction. The Dollonds' ideas were then taken up by an Amsterdam instrument maker called Harmanus van Deijl and he

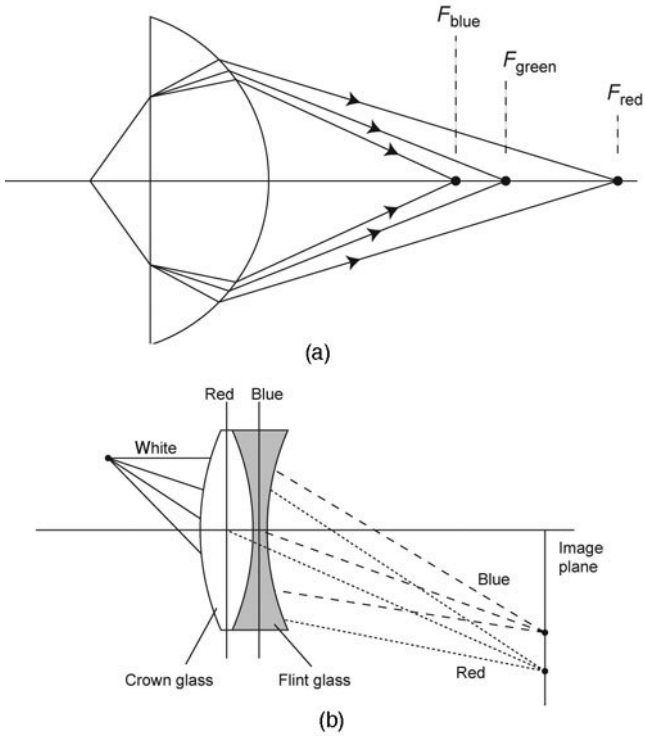


Figure 1.49 (a) The basis of longitudinal chromatic aberration is illustrated. Because of dispersion, see Fig. 1.12(b), different wavelengths are refracted by different amounts as the ray passes through the lens and the best image position is compromised. (b) One solution was to use a pair of lenses – the achromatic doublet – comprising a convex crown glass lens cemented to a flint concave lens. Although this cured the longitudinal chromatic aberration, off-axis sources gave a transverse chromatic aberration effect.

succeeded in correcting the much smaller size microscope lenses for chromatic aberration at the end of the eigthteenth century. Unfortunately, these original achromatic lenses still suffered from spherical aberration (i.e. a curvature of the focal plane) because different lens zones focus at different points along the lens axis, as illustrated in Fig. 1.50.

Around 1820–1830, instrument makers concentrated on making microscopes which did not suffer from spherical aberration. On the one hand, reflecting microscopes gained ground and also use was made of gemstones for compound microscope lenses. However, the major development came in 1830 with the publication in the *Philosophical Transactions of the Royal Society*[11] by Joseph Jackson Lister of a method of using two achromatic lens combinations at a certain distance apart to eliminate spherical aberration from the image. This discovery enabled the optical microscope to become, during the Victorian

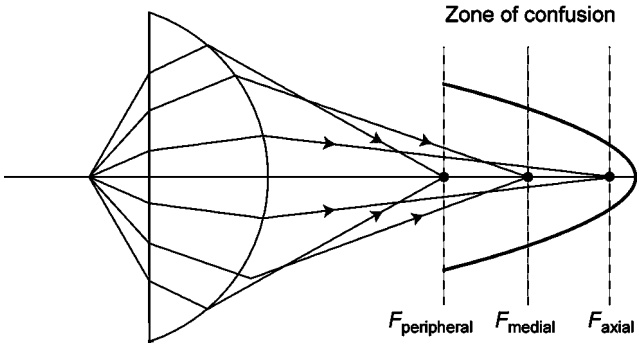


Figure 1.50 Spherical aberration is produced by a single lens because the rays at the extremities of the lens are refracted more than those rays near the axis. By figuring (or modifying) the shape of the lens surface profile, this effect can be reduced, but it is more usual to build an objective from a combination of lenses, which minimise this effect.

period, a vital tool for medical and materials research. Lister's innovation, coupled with the production of achromatic objective lenses, enabled instrument makers (like Andrew Ross, James Smith and Hugh Powell) to move steadily towards the achievement of the highest possible optical resolutions. They improved the rigidity of the optical tube and introduced better coarse and fine focussing. From 1860, binocular tubes were provided for many microscopes (which made extended usage less tiring) and usually the microscopes were sold in large mahogany chests with up to 100 accessories. On the Continent, the German optical instrument trade rose to prominence with designs from Carl Zeiss of Jena and Ernst Leitz of Wetzlar.[12]

Microscope objective lenses have many inscriptions so that the user may verify their exact type. For example, there may be a number '0.17' on the barrel, which refers to the thickness of the cover slip (0.17 mm) for optimum operation, and it will certainly have the numerical aperture (NA) value inscribed. Apart from these two inscriptions, there may be either '160' or '∞' inscribed too. If 160 is on the barrel, it means that 160 mm is the distance from the opening of the nosepiece (where the objective is screwed in) to the top of the observation tube (where the eyepiece should be inserted) and it is where the objective image plane appears. However, if the objective is marked with the symbol '∞', the objective lens combination produces parallel light down the observation tube instead of the focussed light of the '160' objective. The objective is said to be infinity corrected. For proper operation, this type of objective must be followed by a special tube lens, in the light path, which converges the parallel rays to a focus at the plane of the eyepiece diaphragm.

With the discovery of photography in the mid- to late-nineteenth century, the microscope designs had to accommodate the first photomicrographic attachments, which required wide flat-field images. Larger fields of view for

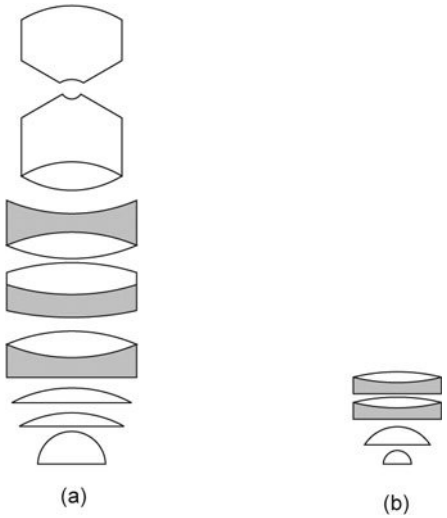


Figure 1.51 The complexity of some modern objective lens systems is illustrated by the schematic diagrams. (a) An oil immersion, plan-apochromat for flat-field work. (b) A simpler (and cheaper) oil immersion achromat. The design brief for an objective lens seeks to find a compromise between minimising aberrations, improving the flat-field properties and reducing the overall cost.

photomicrography required the use of wider microscope tubes of around 30 mm diameter and different types of mount. There are two types of flat-field objectives: planachromats and planapochromats. These lenses are multi-component designs where, for example, an apochromat with magnifying power of $\times 100$ and a numerical aperture of 1.3 would consist of ten component lenses and a similar planapochromat would have fourteen components, as shown in Fig. 1.51. If the objective lens is inscribed 'Planapo', it means that it has been corrected for four colours chromatically and spherically and it is a shortened version of planapochromat (see Fig. 1.52). A variation on this inscription is 'PlanFl' which stands for planfluorite (or Fluars) and these objectives are partially corrected at four wavelengths but not to the same standards as the Planapo types (and hence they are cheaper lenses).

An important consideration in practical microscope designs is the quality of the illuminating system, including the condenser lens. In a research grade microscope, an illuminating system must fulfil three criteria:

1. It must allow maximum resolution and contrast in the image.
2. It should be simple and easy to adjust.
3. The field of view must be uniformly illuminated.

There are two popular systems which satisfy these criteria. Critical illumination, as shown in Fig. 1.53, involves a large area light source *S* and a substage *condenser* lens *C* which forms an image of the source at the object plane, and

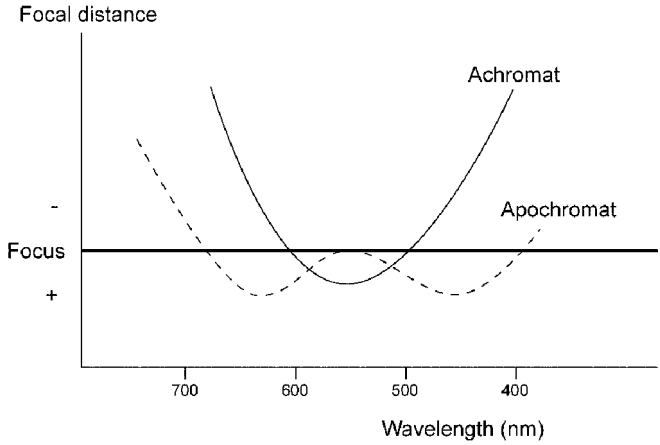


Figure 1.52 The achromat type of objective lens has a focal distance which is a parabolic function of the wavelength. An apochromat objective has significantly reduced chromatic aberrations compared to the achromat and is ideal for colour photomicrography.

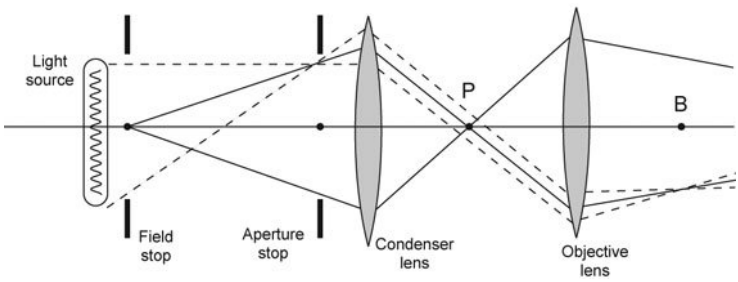


Figure 1.53 Critical illumination is when an image of the light source is formed in the object plane of the microscope by the condenser lens. An image of the aperture stop is also formed in the back focal plane of the objective at B.

which has an NA at least as great as the objective lens in the microscope. Alternatively, Koehler illumination, see Fig. 1.54, can be used (which assumes a small size, high intensity light source is placed at the focus of a well-corrected lamp lens with an extra condenser lens).

Once the appropriate illumination system has been determined, the specimen may have to be viewed directly or by using a cover slip, which is simply a thin glass slide. In biological research a rough surfaced but soft, pliable specimen may be sandwiched between two such glass slides. Hence, high magnification immersion objectives (which only have small depths of field) can better investigate the specimen structure. If an objective lens has been designed for use without a cover slip, it is inscribed with either 'NCG' (no cover glass) or 'NC'.

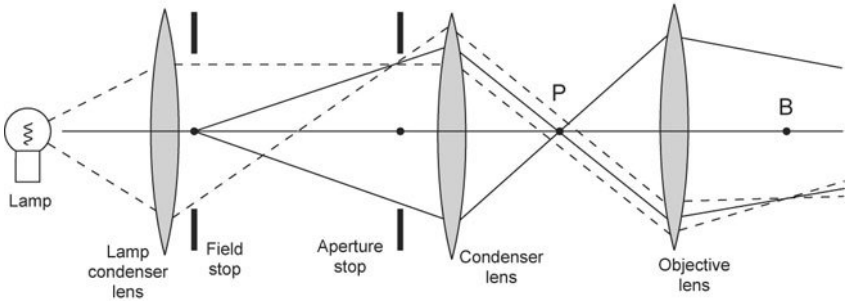


Figure 1.54 Koehler illumination is when the light from a small source is focused by a well-corrected lamp condenser lens onto the aperture stop that is in the front focal plane of the substage condenser lens. This system of illumination enables a small, high intensity, light source to be used because the back surface of the lamp condenser acts as a uniformly illuminated, large area source.

Note that metallurgical objectives are almost always designed for uncovered specimens.

Arguably the most vital component in the microscope is the objective lens. Objective lenses may be 'dry' air objectives or they may be designed for immersion in a coupling medium like oil or water. Usually an objective will be inscribed with 'oil' or 'oel' if it is an oil immersion objective and some manufacturers inscribe a black ring on the lower part of the barrel to indicate that it needs oil contact. Similarly, the letters 'WI' refer to an objective that requires water as the coupling medium. These water immersion objectives are very expensive and, although most useful for biological specimen evaluation, they are of limited use in materials science.

Some objectives may be inscribed with the letters 'UV' in which case they have been specially constructed with glass elements and coatings on these elements to transmit a relatively high percentage of ultraviolet light (wavelengths below 400 nm). These objectives are also useful for near-ultraviolet excitation in reflected light fluorescence studies.

Another important parameter for choosing a suitable objective lens is the working distance of the objective (especially in confocal laser scanning microscopy where it will limit the penetration depth: see Chapter 4). The free working distance of an objective is the vertical distance in mm (or a fraction of a millimetre) from the front of the objective lens to the cover slip or uncovered specimen surface, when the specimen is in focus. A selection of objectives with their working distances are given in Table 1.5.

In most modern microscopes the parfocalising distance (45 mm) is the chosen distance from the opening of the nosepiece to the focussed specimen. By having a standard distance for all objectives, it is easy to rotate each objective, in turn, into the light path with a minimum of refocussing. This set of objective lenses,

Table 1.5 Working distances of different objective lenses

Objective lens type	Magnification M	Numerical aperture (NA)	Working distance (mm)
Eriplan (Air)	50×	0.5	7.00
Eriplan (Air)	100×	0.75	3.00
Achroplan (Water)	40×	0.75	1.98
Achroplan (Water)	63×	0.90	1.45
Achroplan (Water)	100×	1.00	1.00
Planapo (Water) with cover glass + collar	40×	1.20	0.22
Planapo (Oil) with cover glass	100×	1.40	0.10
Planapo (Oil)	60×	1.40	0.17

despite the various lengths of the objectives themselves, is described as being parfocal. (If a focused feature which has been centred in the field of view remains centred as the objective lenses are changed over, the objectives are said to be parcentric.)

Some 'dry' air objectives with magnifications in the range 20× to 60× have built-in correction collars which, when rotated, alter the separation of some of the internal glass elements. These collars can correct for spherical aberration, which occurs when a cover slip is used whose thickness is slightly different from the optimum thickness. In upright microscopes, the range of correction of the collar is from 0.11 to 0.22 mm. For inverted microscopes, the range of correction capability is from zero mm (i.e. uncovered specimen) through to 2 mm (to correct for thick culture vessels).

Finally, in the standard microscope system, there is the eyepiece lens. Eyepiece lenses not only have their magnification inscribed but they are also marked with the numbers 20, 22 or 26.5. These numbers represent the 'field number' of the eyepiece. The diameter of the field of view, in mm, is calculated by dividing the field number of the eyepiece by the magnification of the objective. Hence, if the magnification of an air objective is 10× and the eyepiece has a field number of 22, then the diameter of the field of view would be 2.2 mm.

Eyepieces may also have the inscriptions C or K or WF or H. The letter C or K means it is a compensating eyepiece. (If a microscope objective, like the Olympus LB series, does not include a correction for lateral chromatic aberration, a compensating eyepiece should be used.) WF indicates that it is a wide-field eyepiece. The letter H shows that the eyepiece has a high eyepoint, i.e. the user's eyes do not have to be placed very close to the top lens of the eyepiece, so it is useful to wearers of spectacles. The photo-eyepiece or projection lens, as its name suggests, is a low magnification power eyepiece which is meant to couple the image projected by the objective onto a film plane

inside a camera attached to the microscope. The lower magnification of these eyepieces reduces the problem of empty magnification in the final photomicrograph image.

1.4.4 Optical transmission microscopy

In optical transmission microscopy, the source of illumination transmits light through the specimen and into the objective lens. There are two types of transmission microscope configurations. Brightfield microscopy is when the whole field of view, with no specimen, appears uniformly bright. In a typical brightfield microscope, light passes through the specimen directly into the objective lens. Where the light is absorbed or scattered sideways by parts of the object, the user sees dark areas against a light background. However, in darkfield microscopy, the illumination system is designed so that the light illuminates the specimen from the sides only and no light enters the objective lens when the specimen is absent, as shown in Fig. 1.55. Hence, when parts of the specimen scatter light into the objective, the user sees bright areas against a dark background. Some objectives have a built-in iris diaphragm, which may be partially closed down during darkfield microscopy in order to reduce the numerical aperture of the objective below the lower of the two numerical apertures inscribed on a darkfield condenser. In this way, the darkness of the background may be retained in darkfield observations. For ordinary brightfield

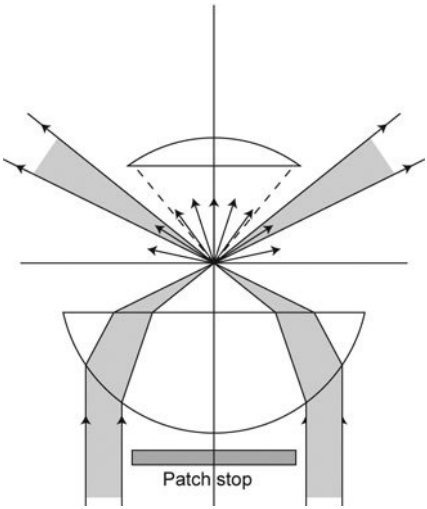


Figure 1.55 For certain objects, it is better to view in darkfield conditions. In this mode of operation, a patch stop below the condenser lens blocks off the direct radiation from the light source. The light rays, which illuminate the sample, can only enter the objective lens and be imaged if there are features which allow large angle scattering of the light.

microscopy, this iris diaphragm is opened fully. However, this mode of operation can only be used in a few special cases in materials science, e.g. when studying thin, transparent films.

1.4.5 Optical reflection microscopy

There are two distinctly different microscope systems that might be termed 'optical reflection microscopes'. The first system uses mainly reflecting surfaces (rather than refractive elements like glass lenses) in order to form a magnified image, as exemplified by the schematic diagram in Fig. 1.56. This design uses catoptric (mirror) objectives and, where lenses have to be used instead of glass, the lenses are fabricated out of materials like quartz which have low attenuation

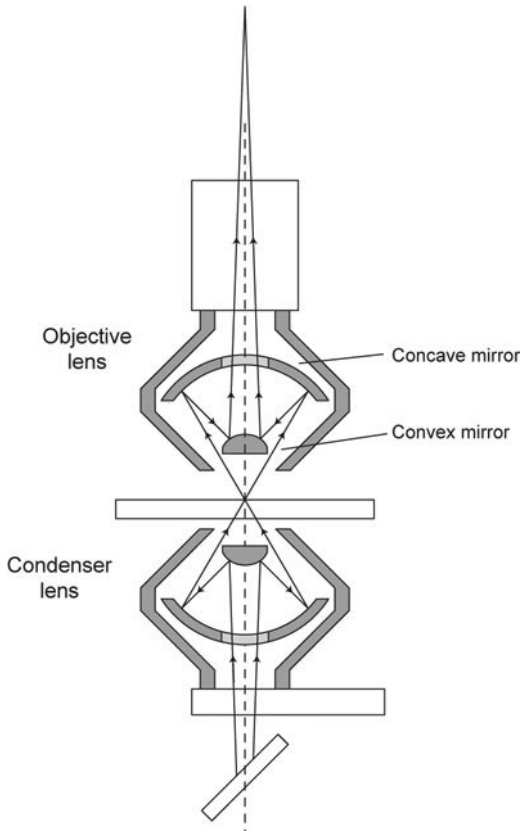


Figure 1.56 A microscope that uses only reflecting optical surfaces is shown. This design is relatively uncommon for optical research, but it becomes important at those wavelengths for which glass attenuates the EM radiation. Hence it can be used effectively for ultraviolet and infrared microscopy.

characteristics at the wavelengths of interest (which are usually either ultraviolet or infrared wavelengths). A microscope designed around reflecting surfaces is particularly good for studying samples over a wider wavelength range than conventional refractive optics. Another useful characteristic is that the reflecting optics microscope design offers even longer objective lens working distances than the lowest magnification lens in refractive designs, as shown in Table 1.5. A reflected light microscope favoured by materials scientists is the Nikon Optiphot where the objective acts like its own condenser. Note the importance of having control over the illumination in these microscopes to reduce glare off the sample surface. Also, they often suffer from a lack of contrast – hence biologists, when they use this type of microscope, tend to stain their specimens.

There are many different variations on a theme of optical reflection microscopy that may be encountered in the research laboratory. These variations include the comparison, metallurgical, darkfield, dissecting, fluorescence, integrating, interference, phase contrast and profile microscopes.[13] Many of these variations are considered in more detail in the following subsections and a standard design of optical reflection microscope system under computer control is covered extensively in Chapter 3.

1.4.6 Fluorescence microscopy

There are two main principles behind fluorescence microscopy:

1. Only those wavelengths that stimulate fluorescence in the specimen are extracted from the excitation light and directed onto the specimen.
2. Only those longer wavelengths, which have been created by the fluorescence effect, (Stokes' Law), are extracted and used for final image formation.

Hence it follows that careful wavelength selection using high-quality optical filters is an essential requirement for fluorescence microscopy. A schematic overview of this type of microscope is shown in Fig. 1.57(a). The principles behind the optical filters – excitation filter and barrier filter – have already been described in this chapter. Note that the intensity of the fluorescence is extremely weak compared to the excitation light intensity and, therefore, to improve the signal-to-noise ratio, the excitation light and the fluorescence must be completely separated (so the optical darkfield illumination technique was used in earlier diascopic fluorescence microscopes). In general, high power excitation light in the near ultraviolet to visible is needed to induce fluorescence and therefore typical light sources are high-pressure mercury lamps or xenon or halogen lamps. Koehler illumination is employed for the latest episcopic fluorescence microscopes and a typical arrangement is shown in Fig. 1.57(b).

Special care is required with this ultraviolet radiation and the optical components must have good transmission characteristics at these wavelengths

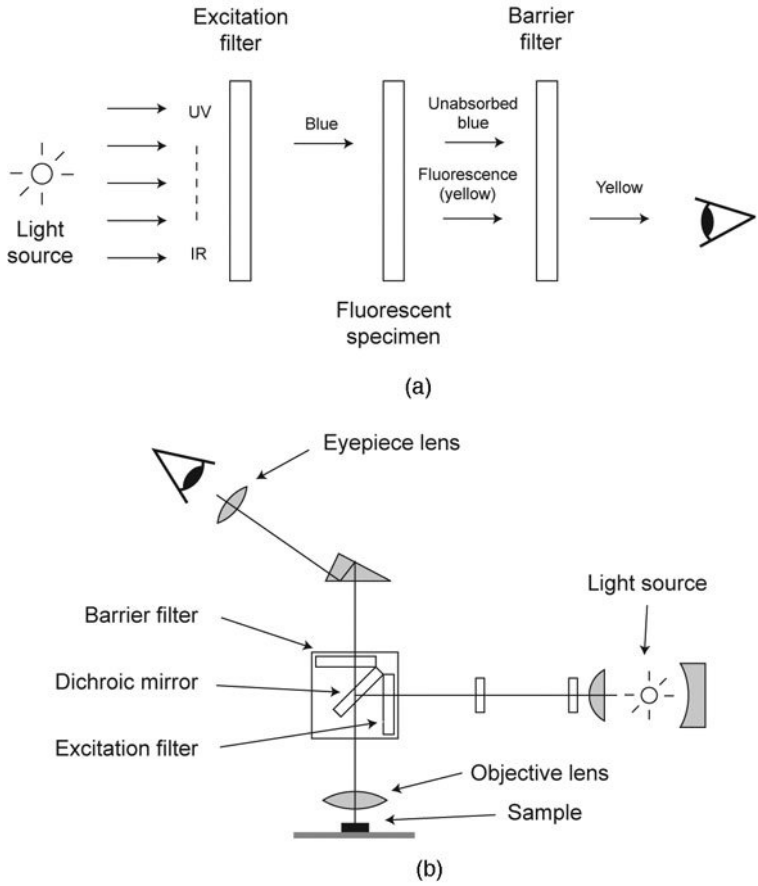


Figure 1.57 (a) The basis of a fluorescence microscope is shown. In this example, the radiation from a broadband source passes through an excitation filter, which transmits a narrow range of wavelengths in the blue part of the spectrum. A barrier filter then separates the longer wavelength fluorescence radiation from residual blue radiation. (b) The most popular form of fluorescence microscope, the episcopic fluorescence microscope, is shown in schematic form.

(and must not fluoresce themselves). Glycerin is one of the immersion liquids used for high NA objective lens work because of its low level of autofluorescence compared to other oils. Another major problem with fluorescence is the photobleaching effect or fading of the fluorescence with time of exposure to the ultraviolet radiation. In order to avoid unnecessary ultraviolet radiation on the specimen during image acquisition, a shutter or neutral density filter is provided at the excitation light source to minimise the UV exposure, which gives acceptable image contrast.

1.4.7 Dark ground, phase contrast and interference microscopy

Some specimens are unable to produce sufficient contrast for interesting regions to be visible under brightfield illumination conditions. However, these regions may produce slight changes in phase in the transmitted or reflected beams. Note that the eye can distinguish differences in light intensity and in colour (i.e. in wavelength) but the eye cannot determine differences in phase. Therefore, in phase contrast microscopy (a special case of interference microscopy), the aim is to translate the variations in phase within the field of view into variations in intensity or colour and hence to make detail visible.

There are two kinds of specimen that could benefit from these techniques: transparent specimens in which there are slight differences in optical density, and opaque reflecting specimens that show variations in surface level. Consider what happens with this kind of specimen. As the numerical aperture of the objective is increased, more diffracted light contributes to the image and the distribution of amplitude and phase across the image plane is a reasonable copy of the amplitude and phase immediately behind the object. If the object is completely transparent (like a thin polymer film) but has an optical thickness that varies within the field of view, it will introduce phase differences in the image plane. The amplitude of the light imaged will be the same over the whole image plane and hence the user will see a uniformly bright field of view. In Fig. 1.58(a), vector lengths represent the light amplitudes (as they have equal lengths this denotes equal light amplitudes) and the different vector directions imply that their phases vary over the field of view. If it could be arranged for a coherent light disturbance to be added to the light at each of these points (see Fig.

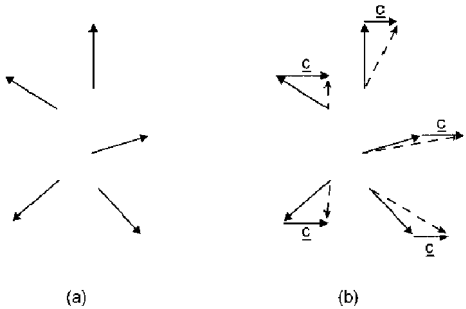


Figure 1.58 (a) In some transparent samples, like thin films, different areas of the sample may vary the phase of the transmitted light, but not the amplitude (which is denoted by the length of each E-field vector). Imaging detectors sense intensity variations, but cannot detect changes of phase. (b) In a phase contrast microscope, the phase changes are converted into intensity changes by arranging for a small fixed phase vector, c, to be added to each of the different phase light vectors. The length of the resulting dotted vectors shows that the amplitude and hence intensity will vary within the field of view.

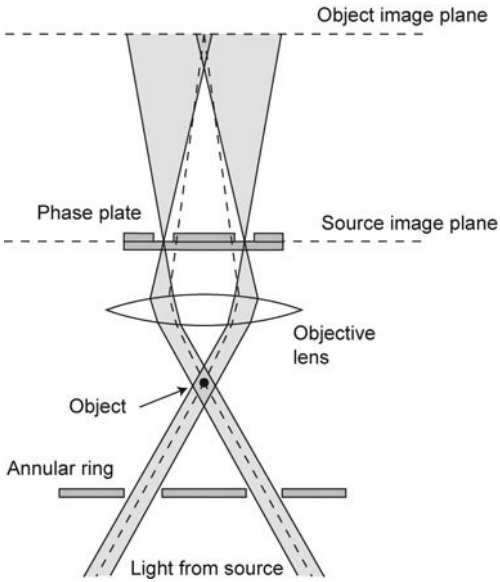


Figure 1.59 A typical phase contrast microscope is shown where the phase plate is configured to add the fixed phase vector and hence convert phase variations into intensity variations.

1.58(b)), the amplitude of the resulting interference pattern would vary from point to point over the image plane. By this means, phase variations are made visible to the microscopist – they are transformed into intensity variations. This is the cunning idea behind the dark ground and phase contrast microscopy techniques and also some of the interference microscope designs.

In the dark ground method, a disturbance is subtracted from every point in the field of view and, in the phase contrast method introduced by Zernike in 1935, part of the light at each point is phase shifted, as shown in Fig. 1.59. Since the process of phase shifting is equivalent to removing it and then adding light of different phase, it is equivalent to adding a constant background disturbance across the image field. A phase plate is a transparent glass plate whose central optical thickness is one quarter wavelength greater than the remainder of the plate. The phase plate is acting as a spatial filter (see Section 1.5.2) affecting the lowest spatial frequencies in the optical beam.

There are two types of interference microscope.

1. A wavefront shearing interferometer where light passing through one part of an object interferes with light that has passed through another part.
2. A wavefront modulation interferometer where an unmodulated reference wavefront is made to interfere with the wavefront which has been modulated by the object structure, as shown in Fig. 1.60.

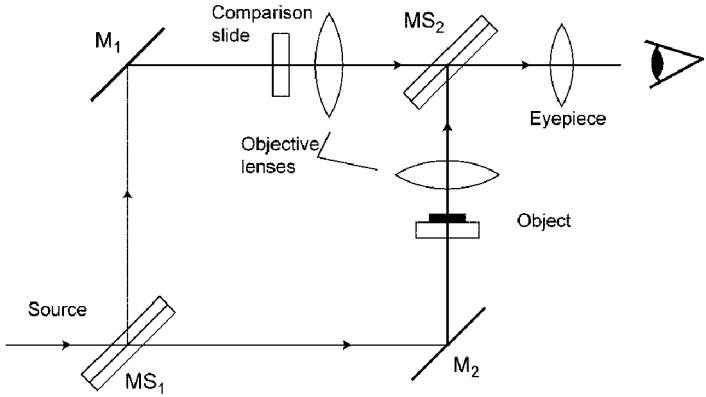


Figure 1.60 One configuration for an interference microscope is the wavefront shearing microscope. The light from a laser source is split by MS_1 and part of the beam is directed through a comparison slide while the other beam is directed through the specimen. The beams are then recombined at beamsplitter MS_2 before detection.

Hence, interference microscopy makes use of the fact that coherent beams of light waves that are out of phase by half a wavelength interfere destructively. Variations in surface levels can be made visible by interference fringe contours or by changes in contrast. Fringe patterns called interferograms are analogous to contour maps and enable accurate quantitative measurements to be made of the surface topography. More recently, surface topography has been determined by alternative novel techniques such as the confocal laser scanning microscopy (which is discussed in detail in Chapter 4).

Another type of interference microscope is the Nomarski interference microscope. A crucial component of this microscope is the Wollaston prism, which is constructed by cementing together two quartz wedges whose polarisation directions are perpendicular, as shown in Fig. 1.61(a). A schematic of the optical components in a Nomarski microscope is shown in Fig. 1.61(b). Note that linearly polarised light from the partially reflecting mirror traverses the prism twice (once before hitting the sample and again after reflection off the sample). For a ray polarised parallel to either direction of polarisation of the prism, the optical thickness of the prism varies continuously from edge to edge. However, because each ray traverses the prism twice, an incident ray passing through the prism a certain distance from the centre, passes through it after reflection at an equal distance from the centre on the opposite side of the prism. Hence, because the total optical path is equal to twice the optical path through the centre, all rays travel the same optical distance through the prism. (This condition is necessary if the rays are to interfere coherently in the image plane.) By rotating the analyser element, the conditions for interference between the two linearly polarised light rays can be varied, leading to certain wavelengths being

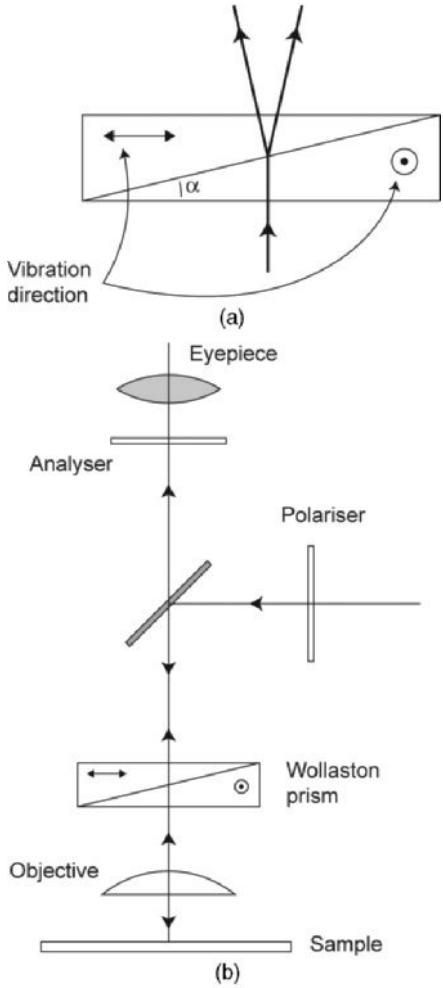


Figure 1.61 (a) A special kind of prism – the Wollaston prism – uses the birefringence effect to split one beam into two beams with different linear polarisations. Birefringent crystals are cut and cemented together as shown to form this prism. (b) The Nomarski interference microscope is a popular design which incorporates the Wollaston prism.

extinguished (and producing a coloured image). When the analyser is set at 45º relative to the polariser, one of the two beams will be attenuated and a normal image will be formed.

Note that objective lenses optimised for use in Nomarski interference microscopes are inscribed 'NIC' and those preferred for another popular interference microscope, the differential interference microscope are inscribed, 'DIC'.

1.4.8 Polarisation microscopy

Yet another way to obtain a sharp image of a transparent specimen is by using crossed polarisers. A polarising microscope therefore differs from standard transmitted light or reflected light microscopes in having polarising filters; the polariser in the illuminating system and the analyser fitted in the microscope tube. One of the filters (at least) must be located in a rotatable, graduated mount. The filters are usually set with their planes of polarisation perpendicular to each other and parallel to the cross-wires in the eyepiece. One or both can be removed from the optical path as necessary, enabling the specimen to be viewed in ordinary or plane polarised light or between crossed polarisers.

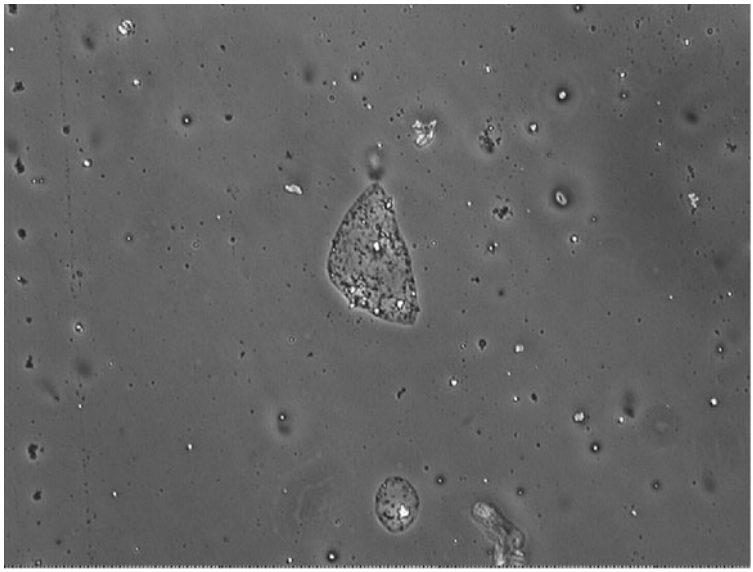
By Malus' Law, when the polarisers are crossed, the field of view will be dark in the absence of any specimen, or if the specimen is isotropic. However, if the specimen is birefringent, the incident plane-polarised light may not be parallel to the specimen's optic axis and, as ordinary rays and extraordinary rays travel at different speeds in these materials, the electric field vector will be rotated. Hence the specimen will appear bright on a dark background.

Accessory plates (at 45° to the planes of polarisation of the crossed polarisers) may be inserted into a slot provided in the microscope tube behind the objective. These accessories help the user to interpret the effects produced by polarised light and they consist of thin wedges or plates of birefringent materials cut in specific ways relative to the crystallographic axes of the materials. A Bertrand lens may be provided for the examination through the eyepiece of interference fringes in the back focal plane of the objective. It is inserted in the optical path of the microscope just beneath the eyepiece. Note that glass and plastics become optically active when placed under stress and the plane of rotation is rotated most where the stress is greatest. Hence, the lenses and cover slips within a polarisation microscope must be strain-free so that the microscope does not introduce any artefacts. However, this phenomenon may also be used effectively by preparing models of bones or machine parts from plastic and observing the model between crossed polarisers to determine the areas of greatest stress.

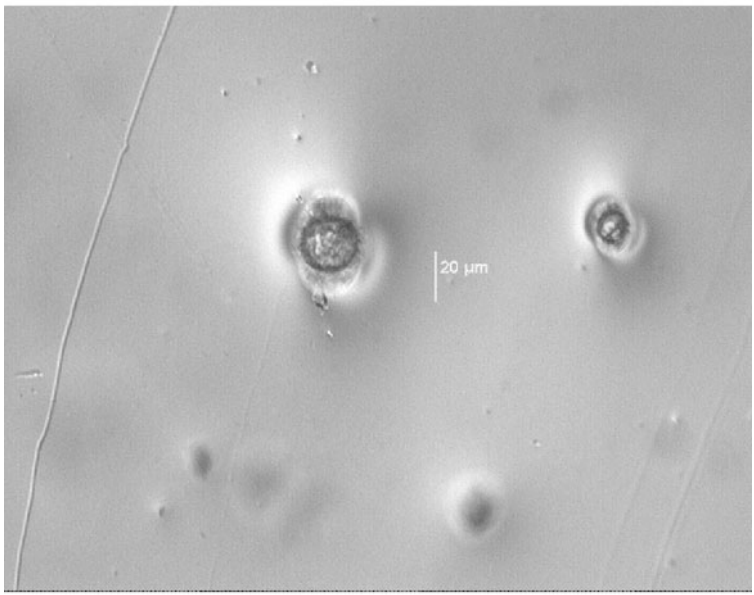
Some typical images from the polarisation microscope and these other types of microscope system are shown in Fig. 1.62 and Fig. 1.63.

1.5 Photonics

So far in this chapter, it has been shown that light – whether or not we regard it as wavelike or discrete photons – can be manipulated in an analogous fashion to electrons. As electrons are the dominant charge carriers within an electronic computer, it is interesting to speculate how an optical computer could be manufactured. Consider the basic operations taking place in an electronic computer and which will be considered in detail in Chapter 2. The signals within a computer are routed from point to point and then modified or transformed.

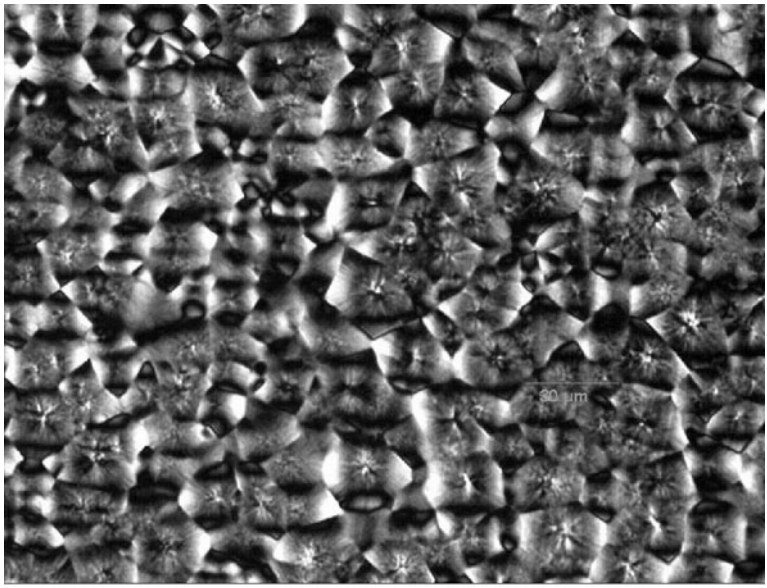


(a)

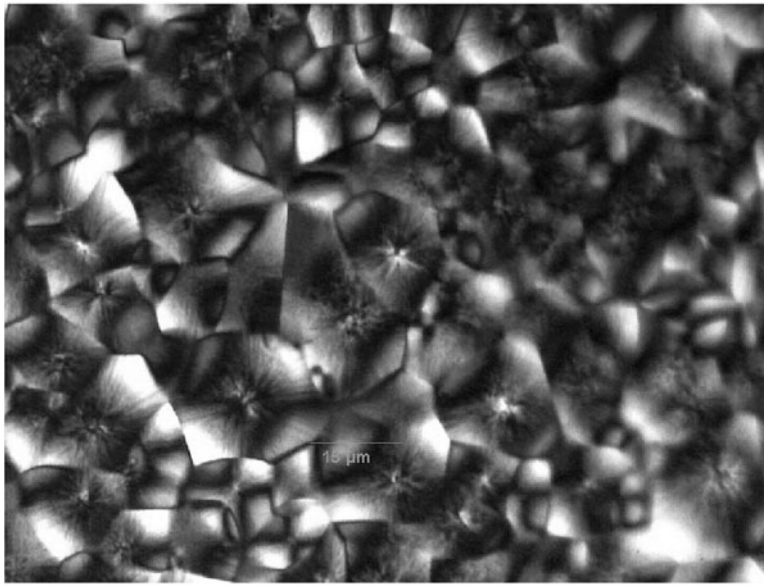


(b)

Figure 1.62 A Nikon Microphot FXA microscope was used to acquire (a) a typical phase contrast image of a PET sample and (b) a typical differential interference contrast (DIC) image of ink.



(a)



(b)

Figure 1.63 (a) and (b) A Nikon Microphot FXA microscope was used to acquire cross-polar images of a sample of red ink at two different magnifications.

Similarly, light beams can be routed from point to point in space (either by a set of mirrors or prisms or, in a more versatile manner, using fibre optics). Light intensity can be attenuated or amplified in the same way that analogue voltage levels are processed, and light can be modulated with broadband signal information. In terms of image acquisition and processing, instead of light levels in a 2D scene being converted by PMTs or CCDs to differing amounts of electronic charge, spatial filters and lenses could be used to transform the 2D image in a highly efficient parallel mode. Also, storage of multiple images can already be achieved in crystalline blocks using holographic techniques. Hence it can be seen that the fundamental activities of the electronic computer (routing, processing and storing) can certainly be mimicked by optical components in large optical bench systems. The study of photonics together with developments in nanometre technology aims to miniaturise these processes so that, ultimately, one might conceive of an optical computer within a single small crystalline block of material![14]

1.5.1 Magneto-optic and electro-optic materials

One important requirement for an optical computer would be the ability to switch light beams very quickly on and off in the same way that binary electronic signals are manipulated. The Kerr electro-optic effect is the double refraction observed when an optically isotropic material is placed in a strong electric field. It is attributed to the alignment of the molecules in the presence of the electric field. The material behaves as if it were an uni-axial crystal in which the electric field defines the optic axis. The relationship between n_{par} and n_{perp} – the indices of refraction in the direction of the electric field and perpendicular to the field – and the applied electric field, E, and the wavelength of the light in vacuum, λ_0, is given by

$$n_{par} - n_{perp} = KE^2\lambda_0 \qquad [1.57]$$

The constant K is called the Kerr constant and high-speed light modulators called Kerr cells have been manufactured, as shown in Fig. 1.64. Also, when certain kinds of birefringent crystals are placed in an electric field, their refractive indices are altered by the presence of the field. This is called the Pockels effect and is found to be proportional to the applied field strength. Fast light shutters (shutter times typically 25 μsec) and modulators are manufactured utilising this effect (see Fig. 1.65).

Another feature of electronic circuitry is its ability for non-linear signal processing, but this too can be mimicked in the optical domain. When intense laser sources of light are used, the high electric radiation field strengths created as the wave passes through an optically transparent medium give rise to a polarising force, P, on the outer electrons within the atoms which is not linearly related to the electric field. This effect results in non-linear performance and

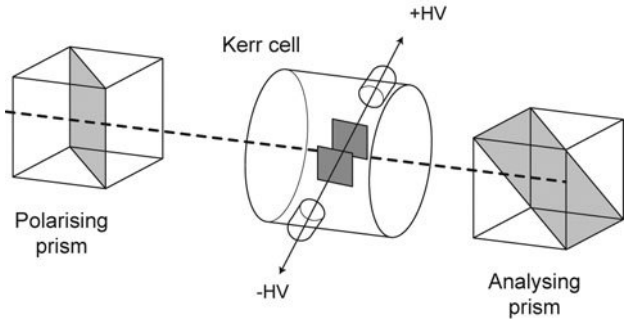


Figure 1.64 The Kerr cell is a fast-acting, optical switch. The first prism polarises the incident light beam. By applying a high potential across the plates, between which the polarised light beam passes, the polarisation of the beam is rotated so that the analysing prism cannot transmit it.

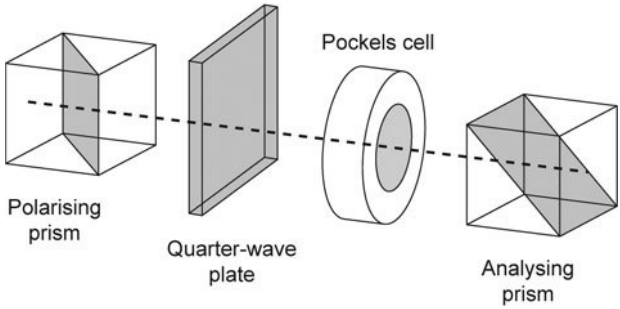


Figure 1.65 The Pockels cell is an alternative design of a fast-acting light shutter and light modulator. It is based on a linear electro-optic effect where an external electric field can alter the refractive index of certain crystals. The Pockels effect is proportional to the applied field strength and the experimental configuration shown is typical for light modulation. The quarter-wave plate provides 'optical bias' and a 50% transmission when no field is present.

hence optical harmonic generation (like frequency doubling), optical rectification and the adding together of different frequencies becomes possible. Once again, the important processes that affect the routing and transforming of signals in electronic circuitry find an optical counterpart in these non-linear effects on the propagation of optical signals through transparent media.

1.5.2 Transformation properties of lenses

Transformation is the key concept when describing the operation of electronic computers. All electronic computer processing entails some form of transformation of input data values to final output values. As will be seen in Chapter 2, the manipulation of 2D images in a conventional computer is computationally intensive. Various convolution filters are applied to the

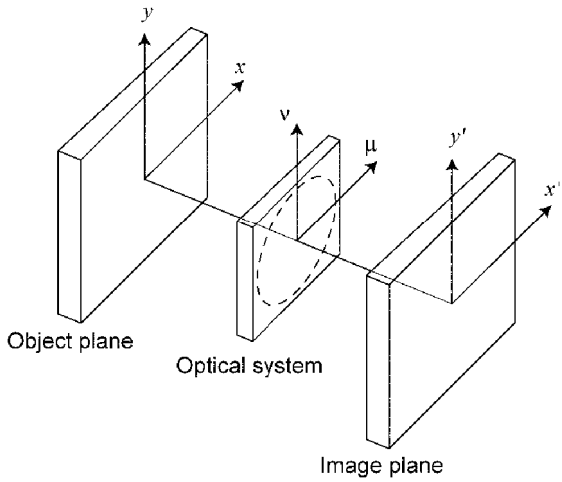


Figure 1.66 Any optical system could be represented by this figure. Features in an object plane are imaged by an optical spatial filter onto an image plane. In practice, each spatial filter has an effective diffracting aperture (which might be the size of the objective lens) and, by using semi-transparent screens, different filtering functions transform the images.

electronic image in order to minimise noise or to detect certain objects within the field of view or to find their Fourier transforms. It would be more efficient if one could keep image data in an optical format whilst the data were being transformed. But what happens when a lens system creates an image?

Consider the geometry of a general diffraction problem in optics, as shown in Fig. 1.66. The diffracting aperture lies in the *xy* plane and the diffraction pattern will appear in the *XY* plane, which is the focal plane of the lens. All rays leaving the diffracting aperture in a given direction, specified by direction cosines α, β, γ are brought to a common focus. This focus is located at point $P(X, Y)$. If a non-uniform aperture is considered, denoted by an aperture function $g(x, y)$, there is a formula which relates this function to the observed pattern at *XY*.

$$U(\mu, \nu) = \int\int g(x, y) e^{i(\mu x + \nu y)} dxdy \qquad [1.58]$$

where the spatial frequencies, $\mu = kX/L$ and $\nu = kY/L$ have been introduced which have the dimensions of reciprocal length (and are equivalent to wavenumber). The functions U and g constitute a Fourier transform pair and the diffraction pattern is actually a Fourier representation of the aperture function.

Now consider an image plane (X, Y) of a coherently illuminated object. The diffraction pattern $U(\mu, \nu)$ of the object function $g(x, y)$ appears in the $\mu\nu$ (or *XY*) plane which is now the optical system. The image function is the Fourier

transform of $U(\mu, \nu)$. If all the spatial frequencies, in the range $\mu = \pm\infty$, $\nu = \pm\infty$, were transmitted equally by the optical system, the image function $g'(x'y')$ would be exactly proportional to the object function $g(x, y)$ and hence a true reproduction of the object. However, the finite size of the aperture at the $\mu\nu$ plane limits the spatial frequencies that are transmitted by the optical system, and with lens defects and aberrations there will be a modification of the function $U(\mu, \nu)$ to a new function $T(\mu, \nu)$ called the 'transfer function' of the optical system.

The transfer function may be modified by placing various semi-transparent screens and apertures in the $\mu\nu$ plane. This is known as spatial filtering. The situation is analogous to the filtering of an electronic signal by a passive network. The object function is the input signal, the image function is the output signal and the optical system acts like the filter that allows certain spatial frequencies to be transmitted but rejects others; see Fig. 1.67.

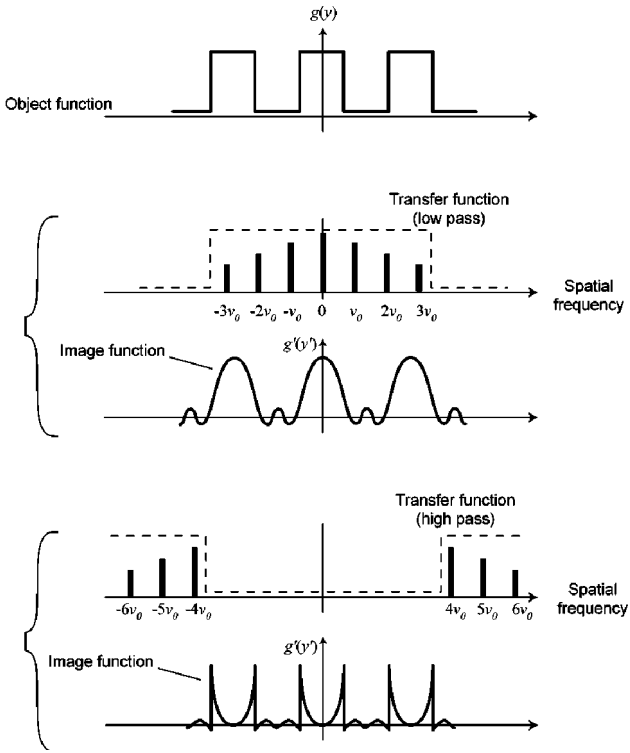


Figure 1.67 Consider the object to be a regular series of on/off gratings. A spatial filter, which performs a low-pass operation by allowing through spatial frequencies between $\pm 3\nu_0$, would give an image function, as shown. If, however, only high frequencies are allowed to pass through the spatial filter, a spikier image response would be produced.

For example, if the aperture in the $\mu\nu$ plane has a width that only allows $\nu < kb/f$, where $2b$ is the physical width of the aperture, it behaves as a low-pass filter. If a central obstruction or screen is placed at the axis of the optical system, the rays on axis will be blocked and this aperture will behave like a high-pass filter. Clearly, fantastic speed-up or throughput could be achieved by such parallel optical computer concepts compared to present-day electronic computers, which have to perform elaborate algorithms on each pixel of the dataset.

1.5.3 Optoelectronic neurocomputing

There is also considerable interest in optoelectronic neural networks which might be trained by example and have adaptable analogue memory as distinct from the binary-tree questioning approach in conventional digital computers. These networks would allow for far greater processing speed in computer vision and personnel interface systems.

For many decades, research has been carried out on electronic equivalents to biological neural nets.[15] The real limitation for large electronic networks where basic processing elements are highly interconnected with each other is the vast amount of copper wiring between elements. The need for parallel interconnects therefore is the driving force behind these optically-based neural networks (since free space optical interconnects allow parallelism due to the non-interactive nature of optical waves). However, another important aspect of workable neural networks is the non-linear processing element and, at the moment, non-linear optical effects are difficult to harness in practice. Nevertheless, it is worth speculating how optical systems might be fabricated in the future.

Most dynamic holographic implementations of neural networks encode the output of one layer of units in the form of plane waves and couple these waves via holographic gratings into plane waves that represent the net input to the next layer of units. The unique encoding between each 2D array of units is carried out via Fourier transforming lenses as shown in Fig. 1.68. The grating structures are formed, as in holographic reconstruction, by passing a coherent expanded beam through an input mask (probably a SLM – spatial light modulator) and lens. At the same time a similar beam (coherent with the first beam) is passed through the holographic material so that this beam recreates the desired output image (this reference beam has the required output plane waves). The beams interfere to form periodic optical intensities within the material, which form the grating structures that allow the output image to be obtained directly from the input image. In this way, the correct interconnections are formed. Holographic materials like silver halides, dichromate gelatine (DCG) and the more recent dynamic photorefractive materials, e.g. iron-doped lithium niobate and barium titanate, allow periodic gratings to be recorded by altering their optical properties, such as transmissivity or refractive index, in response to light intensity. Once formed, each grating can

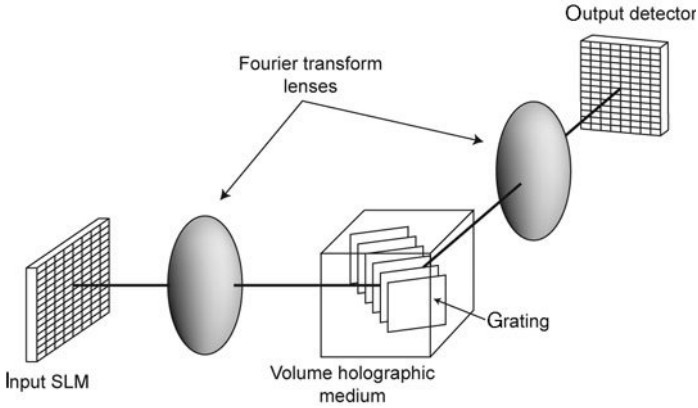


Figure 1.68 An overview of a holographic-based system is shown, which is based on photorefractive materials (able to store vast amounts of image data in a small volume). The image data is stored as variations in either absorption properties or refractive index within a 3D volume.

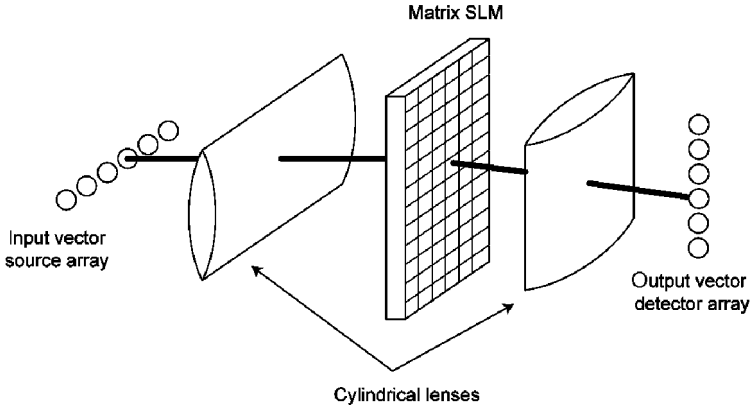


Figure 1.69 This configuration provides a spatially-switched system and can act as a vector matrix multiplier.

be considered an interconnection whose coupling strength, determined by the amplitude of the grating and its physical size, is the weighting. For adaptive neural networks, the holographic material must be dynamic and the gratings are constantly updated after input patterns are presented. One problem area to resolve before practical implementation is the cross-coupling that occurs between the optical waves and gratings in thick specimens.

An alternative method of implementing dynamic connections is to create fixed optical interconnects in the form of optical beams that are spatially separate in one plane, as shown in Fig. 1.69. An array of adaptive modulators

(SLMs), in this plane, modulate or weight the beams. By definition, these systems are 2D interconnect systems where the number of interconnects is proportional to the area of the SLM. The potential size of SLMs in terms of pixels per unit area suggests that medium-size networks with between 100 to 1000 inputs are best implemented with this approach. Above this size, the holographic approach is better suited. The vector matrix multiplier architecture is shown in Fig. 1.69. The optical emission from the input units is expanded with a cylindrical lens in one dimension, passed through a weighting SLM and then converged via another cylindrical lens onto each of the output units in the orthogonal direction. The interconnection between each input and output unit is formed by a unique optical beam passing through a modulating pixel in the SLM. By attenuating these beams, the desired weighting can be achieved. By cascading these optical processing building blocks together, large layered networks can be built for complex processing tasks, incorporating optoelectronic device arrays between the layers to provide the necessary non-linear interfaces. Although this is a 1D operation, refractive or diffractive optical elements can produce expansion or 'fan-out' in 2D for image input and output.

For example, in Fig. 1.70, lenslet arrays (refractive elements) or equivalent computer-generated holograms (diffractive elements) perform fan-out by replicating the input information into an image plane, where input information is spread out into many amplitude-encoded spots. Spatial modulation is then performed which represents the interconnection weightings and fan-in then

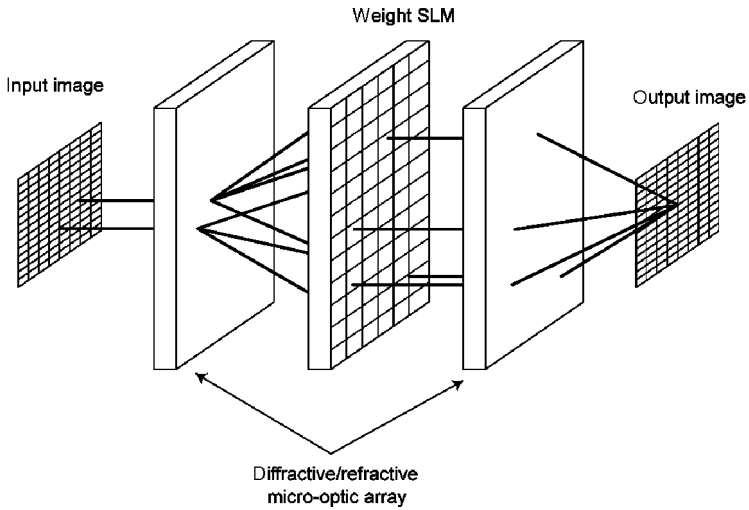


Figure 1.70 A lenslet array is shown and a spatial light modulator (SLM) which consists of many pixel elements. Each pixel element's transmission properties are controllable, so that different transformations can be applied to the images.

occurs, either via summation over a large area detector or further optical elements, before detection. In Fig. 1.70, SLMs are used for the weighting elements and parallel optical in/optical out devices for the optoelectronic interfacing. Although SLMs can be built from GaAs multi-quantum well material, micro-machined silicon or hybrid PLZT on silicon, SLMs based on liquid crystals appear to be best candidates for such devices.

1.5.4 The future for computer-controlled, optical microscopy

Optics is one of the oldest branches of physics and yet is still a very active area for research. In the preceding sections, it has been shown that whatever electrons can do in a computer nowadays, in the near future it should be possible to mimic and surpass with light beams. Many groups (e.g. Snyder[16]) are now researching into optical computer configurations. Although large scale size optical bench arrangements currently illustrate how laser light can be manipulated like electronic signals, a little imagination suggests that these set-ups could be miniaturised significantly. As circuits and devices are miniaturised, an adequate description of the interaction of light and materials will require a quantum optics approach, which was initiated by the work of Paul Dirac.[17]

These photonic devices are important for two reasons. Firstly, they offer the possibility of massive parallelism compared to conventional computers (e.g. in performing Fourier transforms). In addition, they offer ultra-high speed operation because the information carriers will be photons (travelling at the speed of light, 3×10^8 m/sec) rather than electrons whose drift speed through conducting wires in conventional electronic computers is typically 10^{-4} m/sec.

With comparable advances in nanotechnology, perhaps these prototype optical bench systems could be fabricated into a small transparent block of material through which photons digitise and manipulate images under the control of optoelectronic inputs into the side of the block. Recently, the world's smallest microscope has been unveiled,[18] based on the confocal microscope concept (discussed further in Chapter 4). This microscope fits into a physical space of $1\,\text{mm}^3$ and the design is shown in Fig. 1.71. Incredibly, each lens is a tiny drop of polymer in a silicon ring and the ring sits in a silicon platform, which can be moved by electrostriction. In this way, the basic focus and lens alignments can be carried out. A blue light-emitting diode (LED) is the light source and a photosensor detects the light reflected from the specimen. Their next stage is to include a power source for the LED and photosensor and to incorporate a radio frequency transmitter, which could beam images from the microscope to a computer. Clearly, the standard, computer-controlled optical microscope of the not-too-distant future will probably look like, and be the same size as, a sugar cube!

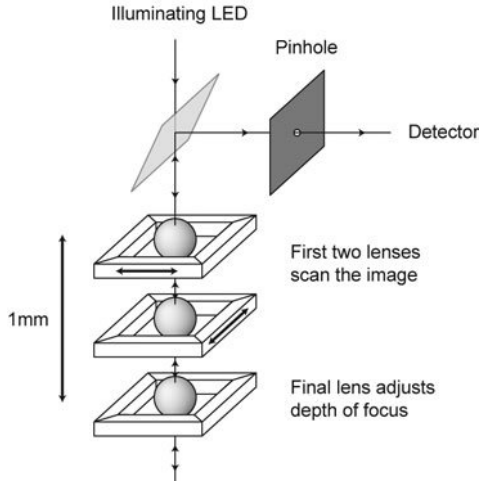


Figure 1.71 A remarkable piece of nanoscale engineering has created recently the smallest confocal microscope packed within a total volume of a few cubic millimetres. A miniature, solid-state, light-emitting diode is the light source and polymer beads form the lens optics. These beads sit on piezoelectric frames and can be positioned in *x*, *y* or *z* to optimise the focusing.

1.6 References

1 B. Greene (2000), *The Elegant Universe*, Vintage.
2 J.C. Maxwell (1873), *A Treatise on Electricity and Magnetism*, 2 vols, Oxford.
3 P.M. Delaney and M.R. Harris (1995), 'Fibreoptics in confocal microscopy', chapter 33 in *Handbook of Biological Confocal Microscopy*, ed. J.B. Pawley, Plenum Press.
4 R.N. Wilson (1996), *Reflecting Telescope Optics* Vol. 1, Springer; (1999) *Reflecting Telescope Optics* Vol. 2, Springer.
5 Project VIRGO, Gravity Wave Detector, website *http://www.virgo.infn.it/central.html*
6 G.N. Watson (1944) *A Treatise on the Theory of Bessel Functions*, 2nd edn, Cambridge University Press.
7 A. Einstein (1905), *Ann. der Physik* (4), **17**, 132.
8 R.B. Lindsay (1970), *Lord Rayleigh – The man and his work*, Pergamon Press.
9 A. Mark-Smith (ed.) (2001) *Alhacen's Theory of Visual Perception: A critical edition*, with English translation and commentary of Alhacen's 'De aspectibus'.
10 R. Hooke (1665), *Micrographia*, **47**.
11 J.J. Lister (1830), On some properties in achromatic object-glasses applicable to the improvement of the microscope. *Phil. Trans of Royal Soc.*, 187–200.
12 G. l'E. Turner (1980), 'Antique Scientific Instruments', Blandford Press Ltd., pp 117–18.
13 M. Herzberger, O.W. Richards and L.A. Harris (1988), 'Optical microscope', in *Optics Source Book*, Science Reference Series, McGraw-Hill, pp 77–90.
14 A. Snyder, 'Light guiding light in the new millenium', Clifford Paterson Lecture to the Royal Society, London, 10 December 2001, reporting on research at the Optical

Sciences Centre, University of Sydney.
15 E. Sanchez-Sinencio and C. Lau (1992), *Artificial Neural Networks: Paradigms, Applications and Hardware Implementations*, IEEE Press.
16 Berkeley Nanoscience Research Center, website *http://www.foundry.lbl.gov*
17 P.A.M. Dirac (1927), 'The quantum theory of the electron', Part I, *Proc. Roy. Soc.* (London), **A117**, 610–24; 'The quantum theory of the electron', Part II, *Proc. Roy. Soc.*, **A118**, 361.
18 C. Zandonella (2001), Sick cells can't escape the eagle eye of the micro-microscope, *New Scientist*, **172** (2320), 24.

1.7 Bibliography

G. L'E. Turner (1980), *Antique Scientific Instruments*, Blandford Press Ltd.
P.A. Tipler (1995), *Physics for Scientists and Engineers*, 3rd edn, Worth Publishers.
R.S. Longhurst (1967), *Geometrical and Physical Optics*, Longman Publishers.
R. Haynes (1984), *Optical Microscopy of Materials*, Blackie & Son Ltd.
S.G. Lipson and H. Lipson (1969), *Optical Physics*, Cambridge University Press.
D.G. Giancoli (1980), *Physics: Principles and Applications*, Prentice/Hall International.
S.P. Parker (ed) (1987), *Optics Source Book*, McGraw Hill.
M. Born and E. Wolf (2001), *Principles of Optics*, 7th expanded edn, Cambridge University Press.
E. Hecht (1975), *Optics*, Schaum Outline Series, McGraw Hill.
B. Hoffmann (1965), *The Strange Story of the Quantum*, Pelican Books.
F.A. Jenkins and H.E. White (1957), *Fundamentals of Optics*, 3rd edn, McGraw Hill.
Spindler and Hoyer (2001), *The Catalogue – Optomechanics GE4*.
G.R. Fowles (1975), *Introduction to Modern Optics*, 2nd edn, Holt, Rinehart & Winston Publishers.
J.B. Pawley (1995), *Handbook of Biological Confocal Microscopy*, 2nd edn, Plenum Press.

2

Digital imaging and processing

> All the great submarine Telegraph Companies, and almost all foreign countries and colonies, have adopted the word-tariff, or system of charging a certain sum for each word, and Great Britain has practically done so by the changes effected during the last Parliament. Every person who has heretofore sent telegraphic messages abroad has learned by experience the economy of condensation, and the advantage of the use of a Code known to both sender and receiver. By this means, the substance of a message embracing a dozen ordinary words may be conveyed in a single code-word, with the fulness and clearness not to be otherwise attained unless at prohibitive cost …
>
> Preface, 'Unicode' – *The Universal Telegraphic Phrase Book*, 9th edn, 1894

2.1 Introduction

The more one thinks about everyday experiences, the clearer it becomes that every individual's experience of the 'real world' is a unique representation of 'what reality is'. The human senses of sight, sound, touch, taste and smell are perfectly adequate for our everyday needs, but they do have their limitations (being accurate only to a certain degree). Our sensors (eyes, ears, fingers, tongue and nose) can have their sensitivities enhanced by external high-tech devices like microscopes, telescopes, hearing aids and robotic arms. However, ultimately, the sensations must be coded into a form that can be processed by our brains. The coding system, being imperfect, will affect our immediate responses to the sensation and the long-term retention of the experience. Incredibly, the retinal image at the back of the eyeball, which is an upside-down form of the true scene, is interpreted by the brain as being the correct way up! The discrete optical nerve impulses reaching the brain are cleverly coded to give the perception of colour, form and brightness information. Human ears can handle a vast range of audio, pressure wave intensities, although we are limited to sensing frequencies below 20 kHz. The human optical coding system can now be fooled by the speed and sophistication of the 3D image reconstruction technology. What truly exists (physical reality) and what has been simulated (virtual reality) is becoming harder to determine. The philosopher Descartes

may have been the first to see man as a machine but it was Turing[1] who first made us think of man as an information processor. It was Turing's ideas that ultimately pointed the way to the von Neumann machine – the first programmable computing engine.

Not only sensual information is acquired by individuals and absorbed by the brain, but also messages, concepts, ideas (in the form of mathematical symbols and text) are continually received/transmitted, stored, routed or modified between individuals and organisations. All information is handled, in effect, by some kind of code. The code of numbers and mathematical operations (addition, subtraction, multiplication, integration) is universally recognised, but text is coded in many different forms, i.e. in each written language (English, French, Mandarin, etc.), the words or pictograms convey a special meaning. The pictograms used by the Chinese are 'word pictures'. In other words, it is possible to have hierarchies of coded information. As an illustration, consider the quotation from the 'Unicode' booklet at the beginning of this chapter. It implies that, to minimise the cost of transmitting a message (and, possibly, to keep that message secret to many people), one coded word could represent a whole phrase or sentence. For example, in the Unicode system, the message 'No money, in great difficulties, send at once £100' was simply the code word 'obeliscus' and the message 'No money, very ill, and want to come home' was 'oblique'!

If one were to devise a general, simple coding system to represent all of these different forms of information or data, it is reasonable to suggest that the simplest coding system should be derivable using a two state code. If the two allowed states in this coding system are represented by the symbols '0' and '1' – which might be in effect a switch being 'on' or 'off', for example, or a short dash and long dash, as in Morse Code – any information can be coded by a string of the '0' and '1' symbols. (Note that a single state coding scheme could be used, but it would be very inefficient because different numbers or text letters could only be represented by a variable length string of the same symbol, say 1s.)

The development of electrical machines and the need for their automation has led directly to the current, two-state structure of computers. In the early days of electrical machinery, manually-operated switches were required to start and stop every electrical function and also these switches were required to select different operational modes, e.g. fast/slow and forward/reverse operations. The invention of reliable clock mechanisms and their increasing sophistication led, in the nineteenth century, to clock mechanisms, which could replace each of these human controlled operations by time-programmed, automatic switching. Also, the concept of feedback mechanisms was firmly established by superb mechanical devices (for example, the regulator mechanisms on clocks and early steam engines). The Victorian mathematician Charles Babbage attempted to build the first calculating engine out of gears and mechanical movements. His design would have worked but the precision required to build the device was too

much for the best engineers of the day. Babbage's visionary ideas were 100 years ahead of his time and it was not until the discipline of electronics had been established that useful electronic calculating engines and hence computers could be developed.

2.2 Digital data

2.2.1 Coded numbers

Our familiar number system is defined with reference to the base 10 (or powers of 10). Any number may be written in terms of the coefficients of powers of 10 and these coefficients are in the range 0 to 9. Any number N can be decomposed with reference to any base B and note that, in general, the coefficients of a number represented in the base B will be in the range (0 to $B-1$).

For example, the usual denary (base 10) number 'one hundred and twenty-three'

$$123 = \mathbf{1} \times 10^2 + \mathbf{2} \times 10^1 + \mathbf{3} \times 10^0 \quad [2.1]$$

could also be represented to the base 5, as follows

$$123 = \mathbf{4} \times 5^2 + \mathbf{4} \times 5^1 + \mathbf{3} \times 5^0 = 443 \quad [2.2]$$

or, in the simplest coding (to the base 2),

$$\begin{aligned} 123 &= \mathbf{1} \times 2^6 + \mathbf{1} \times 2^5 + \mathbf{1} \times 2^4 + \mathbf{1} \times 2^3 + \mathbf{0} \times 2^2 + \mathbf{1} \times 2^1 + \mathbf{1} \times 2^0 \\ &= 1111011 \text{ (called a binary code)} \end{aligned} \quad [2.3]$$

The sequence of binary digits (bits) are normally written most significant bit (highest power coefficient) first (or leftmost bit) and the least significant bit (lowest power coefficient) last (or rightmost bit). A sequence of four bits is often called a nibble and a sequence of 8 bits is called a byte. Longer strings of bits are usually called words. The numbering system to the base 16 is also particularly important in computing, i.e.

$$123 = \mathbf{7} \times 16^1 + \mathbf{11} \times 16^0 \quad [2.4]$$

This is called the hexadecimal numbering scheme and, as the required coefficients are 0 through to 15 in this scheme, the first six letters of the alphabet are used to prevent confusion with the denary numbering system. The coefficients 10 through to 15 are represented by the letters A–F as shown in Table 2.1.

Hence 123 in denary is represented by '7B' in the hexadecimal numbering scheme. The hexadecimal scheme is important in computer programming because memory addresses (defined by 16 or 32 bit numbers) and data values (in 8 bit numbers) within a computer program tend to be expressed in hexadecimal

Table 2.1 The representation of the numbers 10–15 in hexadecimal

A	B	C	D	E	F
10	11	12	13	14	15

codes. Usually, hexadecimal codes (hex codes) are prefixed by the 'dollar' sign, for example

$$\$123 = \mathbf{1}^{*}16^{2} + \mathbf{2}^{*}16^{1} + \mathbf{3}^{*}16^{0} = 256 + 32 + 3 = 291 \text{ in denary.} \quad [2.5]$$

2.2.2 Coded text

Text characters could be coded as a special number in any of these base systems. For example, the letter 'A' could be stored as '$41' (i.e. 65 in denary), 'B' could be stored as '$42' (i.e. 66 in denary), 'C' as '$43' and so on. All other written symbols could also be given an appropriate number code and these are called alphanumeric codes. If the decoding system at the receiving end knows that these numbers are in fact to be interpreted as text characters, rather than arithmetic values, the original message can be retrieved. The alphanumeric characters and their hexadecimal codes are shown in Table 2.2.

2.2.3 Coded processes/operations

The logician Georges Boole unwittingly set the scene for modern digital electronics by defining a self-consistent logical system using Boolean algebra, which is based on the symbols '0' and '1'. In order to make decisions as any human operator would do – to adapt the operation of an electronic control system according to input monitoring conditions – logical (reliable and unambiguous) outcomes must follow recognisable input events. All human decision making can be broken down into a few simple logical operations (for example, logical AND and logical OR). One might decide to switch on an electric light because the ambient light level was low AND one wanted to read in the room. Similarly, an interior car light might need to be activated by a simple electrical circuit if ambient light levels were low AND the driver door OR the passenger door was opened. These are both examples of logical AND and logical OR operations. In Fig. 2.1(a), the basis behind the mechanical switching circuits which could perform these operations is illustrated. The logical AND function is represented by two switches in series with each other and the logical OR function is represented by two switches in parallel. In practice, real switching applications require complex logical operations to be performed. Boolean algebra allows the engineer to simplify such circuits or find equivalent representations of combinations of logic gates to solve the problem.

Table 2.2 The matrix of alphanumeric characters (ASCII)

ASCII	Char	Signal	ASCII	Char	ASCII	Char	ASCII	Char
0		NUL	32	Space	64	@	96	‘
1	☺	SOH	33	!	65	A	97	a
2	☻	STX	34	”	66	B	98	b
3	♥	ETX	35	#	67	C	99	c
4	♦	EOT	36	$	68	D	100	d
5	♣	ENQ	37	%	69	E	101	e
6	♠	ACK	38	&	70	F	102	f
7	•	BEL	39	’	71	G	103	g
8	–	BS	40	(	72	H	104	h
9	○	HT	41	)	73	I	105	i
10	–	LF	42	*	74	J	106	j
11	–	VT	43	+	75	K	107	k
12	–	FF	44	,	76	L	108	l
13	–	CR	45	–	77	M	109	m
14	–	SO	46	.	78	N	110	n
15	☼	SI	47	/	79	O	111	o
16	➢	DLE	48	0	80	P	112	p
17	➢	DC1	49	1	81	Q	113	q
18	⇕	DC2	50	2	82	R	114	r
19	!	DC3	51	3	83	S	115	s
20	¶	DC4	52	4	84	T	116	t
21	§	MAK	53	5	85	U	117	u
22	■	SYN	54	6	86	V	118	v
23	↑	ETB	55	7	87	W	119	w
24	↑	CAN	56	8	88	X	120	x
25	↓	EM	57	9	89	Y	121	y
26	→	SUB	58	:	90	Z	122	z
27	←	ESC	59	;	91	[	123	{
28	–	FS	60	<	92	\	124	|
29	↔	GS	61	=	93	]	125	}
30	▲	RS	62	>	94	^	126	~
31	▼	US	63	?	95	_	127	–

The most basic logical operations which underpin any process are given by the truth tables shown in Table 2.3.

The mechanical door switches shown in Fig. 2.1(a) are most likely micro-switches which open or complete a circuit and hence create well-defined voltage levels at the output of that circuit after either door has been opened or closed. This is a key idea: the presence of a positive voltage represents a switch being closed (say) and zero volts represents the switch being open. If this convention is used, a complex mechanical switching network can be simulated by electronic circuits and hence the switching or routing of data can be accomplished at very high speeds. Another key idea is that, for an electronic circuit to automate any human activity, there must be some way to replace the human judgement which initiated the action! In most control systems, this is achieved by using a sensor and a thresholding

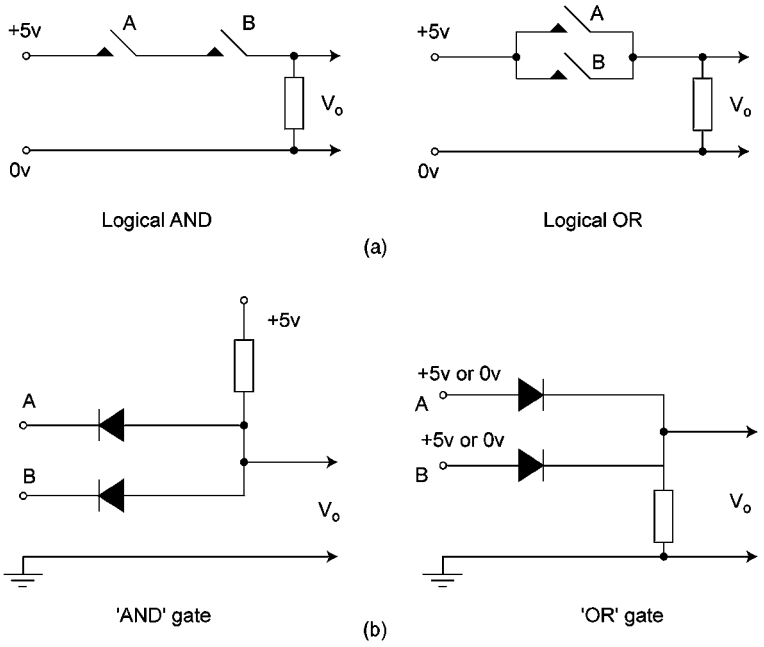
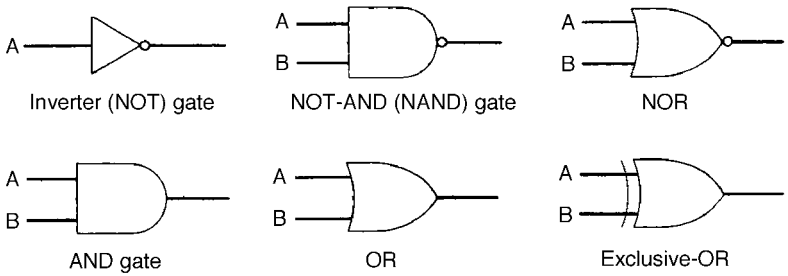


Figure 2.1 (a) Manual switches A and B give an output voltage, $V_o = +5$ volts when both are operated (logical AND) or when either switch is operated (logical OR). (b) The semiconductor diodes are oriented so that the output voltage, $V_o = +5$ volts when inputs A and B are both at $+5$ volts (logical AND) or when either input is at $+5$ volts (logical OR).

Table 2.3 OR, AND, NOR, NAND, EX-OR, EX-NOR truth tables, symbols and gates



A	B	AND	NAND	OR	NOR	EX-OR	EX-NOR
0	0	0	1	0	1	0	1
0	1	0	1	1	0	1	0
1	0	0	1	1	0	1	0
1	1	1	0	1	0	0	1

circuit. Sensors are often special types of resistor elements whose resistance changes when a physical parameter (like pressure, temperature or ambient light level) alters. Consider a sensor (say a light-dependent resistor) connected in series with a fixed resistor and then to a battery, or other power source, to form a potential divider. Variations in light intensity will cause correlated variations in the potential across the sensor. A thresholding circuit compares the potential across the sensor to a fixed potential and, if the sensor potential is higher than the fixed potential, it gives a positive voltage at its output terminal. If the sensor potential is lower than the fixed potential, then the output of the thresholding circuit will be zero volts. It can readily be seen that, if the threshold potential is set to that level of ambient light intensity which the user judges to be inadequate for reading a book comfortably, then the electronic circuit can act independently. In effect, the circuit is carrying out the wishes of its human creator and is retaining a previous human judgement value. Note that the sensor, together with the thresholding circuit, has made the simplest decision: namely, a light bulb must be switched ON (or not) and this output logical state (positive voltage or zero volts) will be combined with other logical outputs to operate (or not) the light bulb. This is the basis of conventional digital logic but, since the 1960s, another branch of logic has been developed by Zadeh[2] called 'fuzzy logic', which attempts to mimic more closely the decisions that humans make. When humans exercise control, their actions most often are the result of 'estimations' about the environment to be controlled, rather than precise operations based on accurate input data. Despite this state of affairs, we are quite good at driving cars and piloting aircraft and fuzzy logic addresses these issues of controlling with fuzzy data.

2.3 The history of digital computing

2.3.1 The transistor age

During the Second World War, Turing and his colleagues at Station X in Bletchley Park designed and built the first thermionic valve-based computing engine to unravel the German Enigma codes. The theoreticians had the help of Post Office engineers with experience of large-scale switching networks. After the war, the possibilities for large-scale computing were appreciated in the UK and the USA and Shannon and Weaver[3] developed 'Information Theory' to describe the routing of information through telephone exchanges. At the same time, scientists and engineers developed large, valve-based forerunners of today's computing machines. When the transistor amplifier had been developed, the miniaturisation and significant power reduction implications could be harnessed to initiate the ever-increasing computer power/unit volume spiral which continues today.

In Fig. 2.1(b) the equivalent semiconductor diode circuits are shown which implement the same truth tables as the mechanical switch circuits. The problem with these diode circuits is that, as more circuits are connected to form a large

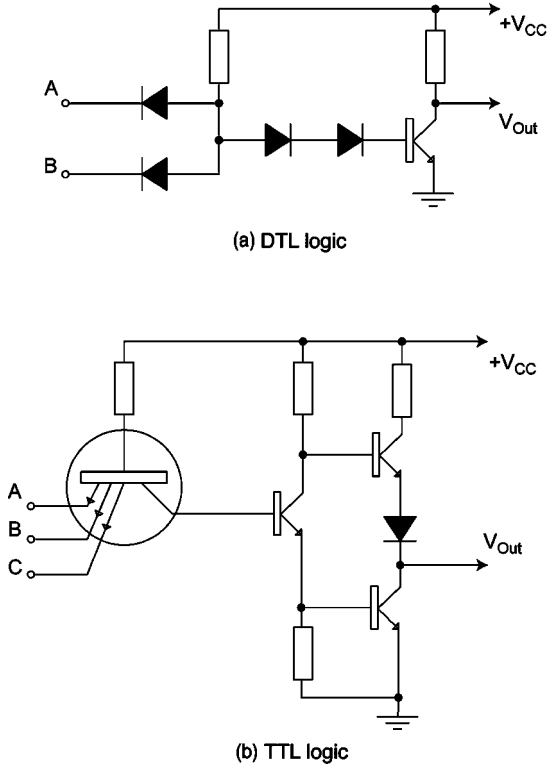


Figure 2.2 (a) A combination of diodes and transistors, called diode-transistor logic (DTL), improves noise suppression. The two diodes in series ensured that the transistor was cut off when either input (A or B) was at zero volts. (b) In the 1970s, more complex multi-emitter transistors were incorporated into logic gates, and the integrated circuits were called transistor-transistor logic (TTL). Both circuits shown here implemented the NOT AND (or NAND) logic function.

network, the voltage levels are compromised due to the loading of one circuit by another. Hence, logic circuits with amplification are required to improve the signal-to-noise ratios throughout the whole logic network, thereby making the circuit more reliable. The bipolar and multiple-emitter transistors fulfilled that need (see Fig. 2.2), and the eventual development of the field-effect transistor (FET) led the way to the fabrication of many transistor cells on a silicon chip. This was achieved by successively depositing different doped semiconductor layers and insulating oxide layers to create the transistor action.

In order to function properly – to think, to make decisions – humans need to refer to their knowledge – background memory – and perform operations based upon the time-order of input stimuli. The logic circuits (or logic gates) discussed so far produce their logical output irrespective of the time order in which the

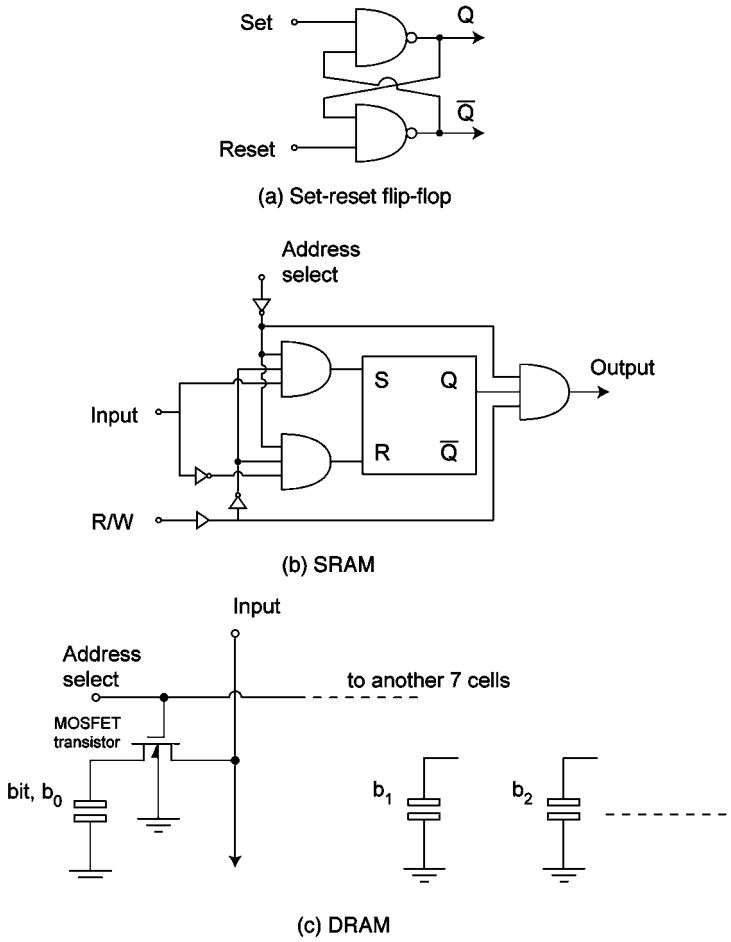


Figure 2.3 All computers contain large numbers of electronic memory cells. (a) The set-reset flip-flop is the basis of most memory designs. It has two stable states and flips between these states when a negative-going pulse (from logic 1 to logic 0) is input on either S or R. (b) A static random access memory cell design is shown (SRAM). The logic state on the Read/Write (R/W) control line determines whether data is to be output from memory or input into memory. The address select control line enables or disables the cell. The multi-input AND gates ensure correct operation. (c) A dynamic RAM (DRAM) design is shown where the charge on a capacitor determines the state of a bit of memory. The charge is controlled by a metal-oxide-silicon, field effect transistor (MOSFET). These FET transistors are used in place of TTL in low power applications.

stimuli occur. Human memory stores information which can be retrieved and which can be modified in some way. Given that we do not know exactly how the brain memorises information (and succeeds in recalling it when required or triggered by sensory inputs), it might seem perverse to attempt to build an electronic memory. However, the presence of a voltage, V, can be represented as

a charge, Q, on a capacitor, C, where $V = Q/C$ and hence a sequence of binary digits (bits) can be represented by a series of charged or discharged capacitors. Unfortunately, charge will eventually leak away from any capacitor (and so ancillary 'refresh' circuitry may be required with this form of electronic memory – as in dynamic random access memory (DRAM)). The memory chip is described as having 'random access' because any of its memory locations can be accessed in any time order (not just a set sequence, i.e. the first, then the second and then the third memory location, etc.).

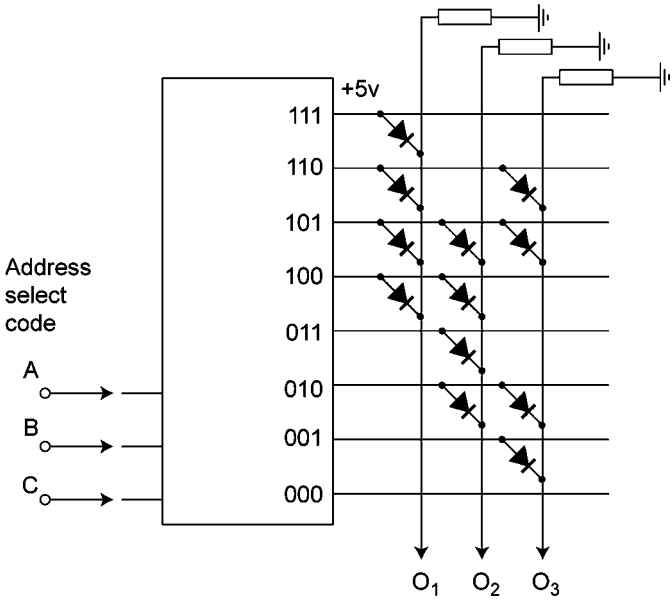
Another versatile circuit design was needed which could flip into one of two states when a short event (for instance, an incoming voltage pulse) occurred. This circuit must maintain that state until it could be changed into the opposite logic state by another short event (or incoming pulse). A typical set-reset flip-flop circuit, which accomplishes this task, is shown in Fig. 2.3(a). This simple latching circuit is the basis of more complex and versatile static random access memory (SRAM) chips within modern computers. For example, all memory chips are organised so that they have 8-bit (one byte) data at a specific memory address. The memory address number may also be coded as a 10-bit, 12-bit, 14-bit or 16-bit value. In order to access the data, the appropriate address code is routed to the input address pins on the memory chip. In Fig. 2.3(b) the logic state on a control input line (the R/W line) determines the direction of data flow. If, say, the control line is at logic 1, new data (appearing at the data pins) will be written in that memory location. If the control line is at logic 0, old data will be read from that memory location and sent to the data pins on the memory chip for transmission elsewhere. In Fig. 2.3(c), the basis of the dynamic RAM is shown for comparison.

Another form of memory that is vital for any computer-controlled instrument is memory that is unaffected by removal of the power to the computer. In other words, memory is required that contains the instruction codes, which enable the system to initialise itself when the power is turned on. This memory is called read only memory or ROM, as shown in Fig. 2.4.

The latest memory devices have ever more sophisticated circuitry which save the user the need for extra logic chips in specific applications. The first type of multi-purpose memory was the dual-port, random access memory which allowed the user to read from one memory location whilst writing to another memory location in the same chip. Hence one could alter the intensity information in one pixel element while sending out other pixel values to be displayed. Further integration has been achieved with the video RAM (VRAM) designs (see Fig. 2.5). These allow the user to modify the memory elements at the same time as a serial, composite video signal is being output to a monitor or computer.

2.3.2 Historical development of microprocessors

The invention of the transistor in 1948 by the physicists Shockley, Brattain and Bardeen, who were studying semiconductor surfaces at the time, has been well documented.[4] Their 'point contact' device was ultimately superseded by



(a)

A	B	C	O_1	O_2	O_3
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	1
0	1	1	0	1	0
1	0	0	1	1	0
1	0	1	1	1	1
1	1	0	1	0	1
1	1	1	1	0	0

(b)

Figure 2.4 Read-only memory (ROM) is non-volatile memory, i.e. even when the power is interrupted, the memory codes remain the same. One design incorporates an address selector and a diode network. (a) Considering the simplified 3-bit circuit, when address lines A, B and C are all at logic 1, the +5 volts on the data input line is routed to the top diode (all other outputs are at zero volts). Current flows through the top diode and the fixed resistor to ground. Hence, the output O_1 will be at a voltage close to +5 volts (logic 1 state) and the other outputs, O_2 and O_3 will be at zero volts. (b) The output logic states are specified for all eight possible combinations of the three address inputs A, B and C. Electrically eraseable ROM (EEROM) is popular because of its ability to be easily reprogrammed in electronic development systems – a special control voltage must be applied to alter the codes.

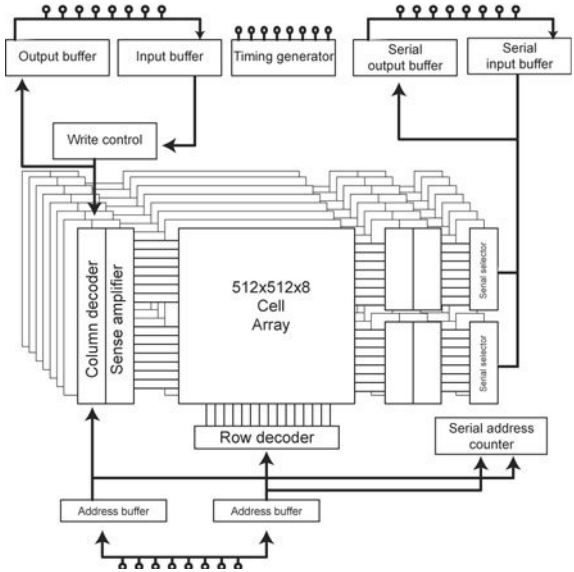


Figure 2.5 In the latest framegrabbers, video RAM chips of incredible complexity are used. In these memory integrated circuits, like the Toshiba TC528267, there are 512×512×8 bits of memory and new bytes of image data may be entered while existing bytes of data in memory are output for display or for further processing.

Shockley's later discovery of the bipolar junction transistor. Shockley also discovered the field-effect transistor, which led directly to the development of the basic building block of modern electronics – the integrated circuit on a silicon chip. The integrated circuit was first described by Kilby and Noyce as 'one piece of solid into which have been included several components, passive as well as active, without external connection between these devices'.[5] The introduction of these microchips in 1959 by both Texas Instruments and Fairchild Semiconductor Corporation marks the start of the modern electronic revolution. As the manufacturing of the monolithic integrated circuits improved, greater miniaturisation was achieved until, in the late 1960s, the first microcomputers incorporating the microprocessor chips came into commercial production.

2.3.3 Motorola 6802

The Motorola 6802 was a very popular 8-bit microprocessor chip. In order to try to understand the operation of a microprocessor, it is worthwhile noting the close analogy between the overall structure of a microprocessor system and the structure of a typical company or organisation. In any organisation, the management controls the flow of information (data), i.e. data has to be sent to, and retrieved from, the company archives. Also, company decisions are made by

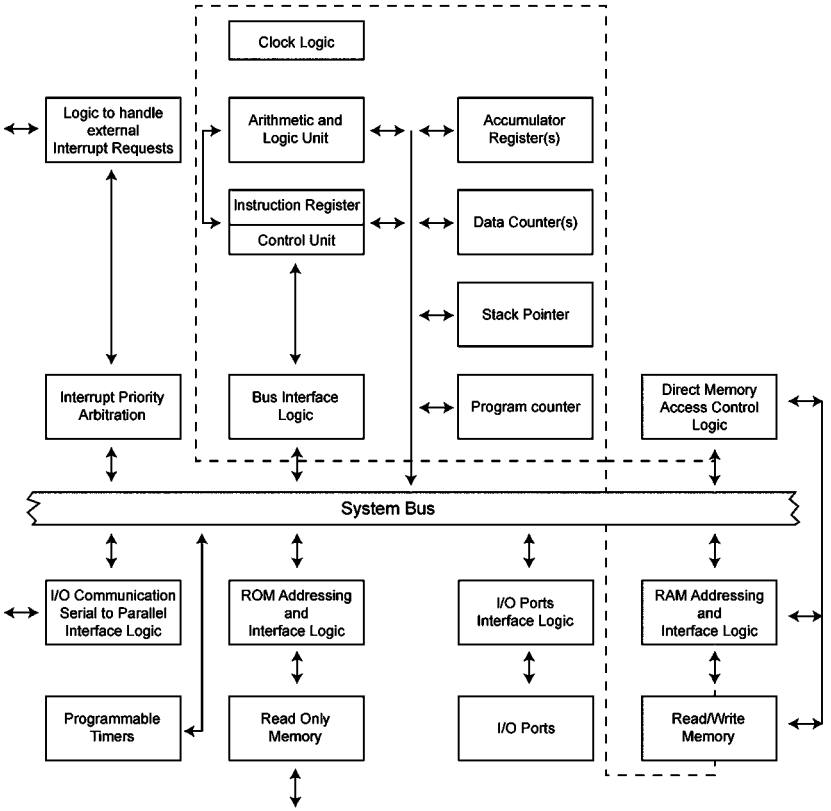


Figure 2.6 A typical microprocessor design from the 1980s was the Motorola 6802 chip, which was an 8-bit central processor unit. Some of the special registers in this chip were the program counter (which kept a record of the memory address of the next executable instruction in the program) and accumulators A and B were two registers where data bytes could be manipulated, added or subtracted.

management on the basis of that data (under a general set of company rules). Similarly, the Central Processor Unit (CPU) (see Fig. 2.6) supervises the execution of a program (usually held in non-volatile memory or read only memory). During the execution of that program, data will be read from or written to volatile memory or random access memory (RAM) after the appropriate routing and elementary processing operations have been performed within the CPU.

The CPU was connected to these memory elements via two distinct multi-copper tracks on the printed circuit board, which were called ‘busses’. The earliest devices had a 16-bit address bus and the 16-bit code on those tracks defined the memory address to which, or from which, an 8-bit datum value was moved. The 8-bit data bus contained the code for the datum value which was destined for the memory location (or had come from the memory location) defined by the address bus code. Not shown on the simplified diagram in Fig.

2.6 were also a few control lines which, for example, defined the read from, or write to, function of the data transfer. One important point is that only one memory chip, or other device, connected to these busses could actively communicate with the CPU at any one time. Hence, the algorithm running on the system had to perform operations in a serial manner.

In the same way that operations by workers within a company are ruled by real time clocks, all operations within the CPU and the associated memory elements are synchronised to the changes of state from a crystal-controlled clock. The basic operations generally require a few clock cycles to implement and so in the 1970s, when the clock frequencies were around 1 MHz, the typical time to execute an operation was of the order of a few microseconds. Therefore these early microprocessors could perform, typically, around 250,000 to 300,000 operations per second.

Another important company activity is marketing and publicity (where it deals with the outside world). These activities are paralleled by one, or more, input/output chip(s) in a microprocessor system. These chips were variously called peripheral interface adaptors (PIAs), peripheral input/output (PIOs) or versatile interface adaptors (VIAs). They were given a number of specific 16-bit addresses and some of these addresses located 8-bit codes, which were to be transmitted to another device or inputted from another device. Other memory locations contained control codes which could be manipulated by the program running on the CPU. The fastest method of communicating bytes of data was through a parallel communication mode (where 8-bits at a time were placed on an output connector and a synchronised handshake took place between sender and receiver). Otherwise data were sent out and received via a serial link where each bit was sent out (or received) in sequence along one single line according to an accepted protocol like the RS232 protocol. The serial connection was slower to transmit data but it required fewer connecting lines (only two lines – signal and ground).

In the 1970s microprocessor clock speeds were around a few megahertz and hence their basic ‘fetch and execute’ cycles to complete a particular instruction or operation would take a few microseconds. At this time, there was considerable interest in datalogging using microprocessors and computer systems. The idea behind the datalogger is that analogue signals (from many sensors) are digitised at regular intervals by analogue-to-digital converters (ADCs) and the digital values are saved in memory. Economical 8-bit ADCs at this time took 15–20 microseconds to digitise the analogue signals, but higher resolution (16-bit resolution) and faster ADCs (like the flash ADC discussed later for video work) could also be purchased.

2.3.4 Vela and Labview – instrument simulations

All measurement systems have similar characteristics. The system must have an input sensor, which converts variations in the physical parameter of interest to

voltage variations. These voltage variations may be stored so that, instead of single discrete samples being displayed in analogue or digital form, the mean value of a number of measurements may be computed and displayed. (This technique is common in order to minimise the random noise from the sensor.) In the mid-twentieth century, single function measuring instruments like phase meters, counter/timers, power meters, capacitance meters, etc., were marketed. However, with the advent of the datalogger concept, it became possible to simulate many instruments by software. A datalogger is a device, which is microprocessor-controlled. It uses an ADC to digitise each input waveform, stores the digital output of the ADC in RAM memory and may process the data before outputting. The datalogger outputs the digital data either as an analogue signal (through a digital-to-analogue converter, DAC) or digital values with seven-segment displays or on a liquid crystal display screen. One economical multi-purpose laboratory aid for school science laboratories, which incorporated the Motorola 6802 microprocessor, was called VELA.[6] Over 80 machine code programs within four EPROMs could be accessed to simulate different types of measuring device (including all of the single function instruments mentioned above). An overview of the main integrated circuit chips within the VELA is shown in Fig. 2.7. The only drawback to this simulation approach is the slower speed of the software simulation process compared to a dedicated, hardware-driven, single function instrument (which in many applications may not be a problem at all). More recently, National Instruments have been working on their LabView system. In LabView, the user can select from a suite of virtual measurement devices, interconnect these virtual instruments on the computer screen and define what range of measurements are to be taken.[7] The computer has a special input/output card which communicates with the appropriate sensors embedded in the particular experiment and which allows these measurements to proceed.

2.3.5 The link between software language and the electronics

In the 1980s the microprocessor clock speed increased and the microprocessor chips became more complex in terms of their range of instructions or op-codes that the user could select. To illustrate the link between the logic gate circuitry and the different levels of the Motorola 6802 microprocessor software programming, consider the situation shown in Fig. 2.8. An 8-bit code has to be routed from a memory address $1234 into one of the special CPU registers Accumulator A or Accumulator B. (This action would be the first stage in the process of adding two numbers together.) The assembly language expression or line in the program would be either LDAA $1234 or LDAB $1234. These expressions represent a three-byte instruction in the machine code, which the 6802 CPU understands. These three bytes would be in three contiguous memory locations within the EPROM storing the program. The first byte of the operation code is either B6 or F6, which sets the internal logic gate switches to achieve the

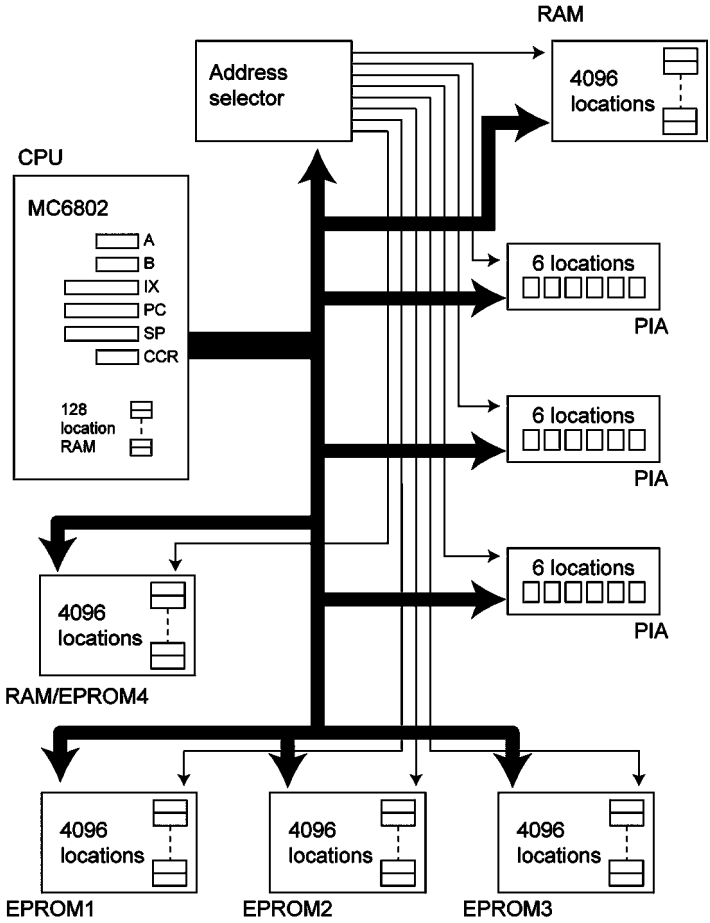


Figure 2.7 An overview of the VELA device, which was a microprocessor-controlled, multi-functional instrument for science laboratories. In software, it mimicked different simple measurement devices, e.g. timer, frequency meter, datalogger, counter, phase meter, etc. Over 60 different measurement functions were coded in the three EPROMs. One of the peripheral interface adaptors (PIAs) was used to enter instructions via a keypad and display data to a row of seven-segment displays. Another PIA interfaced with preamplifiers, an ADC and a DAC and the third PIA enabled parallel expansion and communication to printers or other digital devices.

appropriate routing. The last two bytes represent the address in memory where the data value is to be found (most significant byte first). The only difference between the two op-codes is the logical state of one of the bits in the operation code. Hence, one could imagine that the logic status of this bit would control the routing of each line of the 8-bit data value from memory address $1234, as shown in Fig. 2.8.

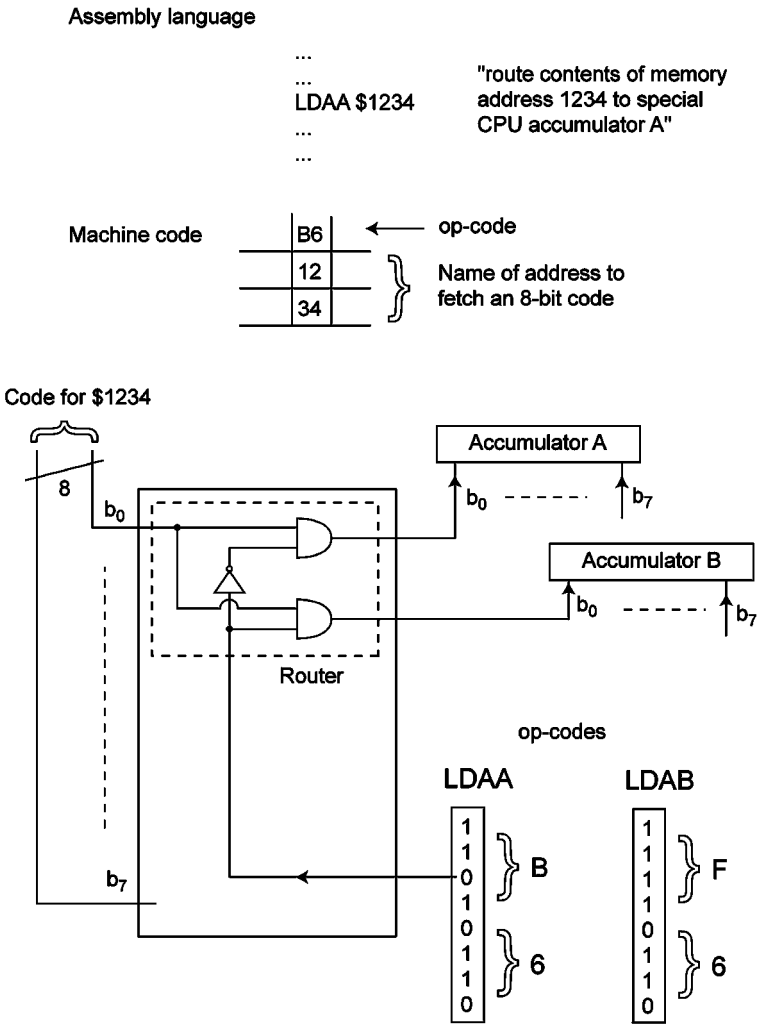


Figure 2.8 In order to make the connection between the software resident in the Vela and the hard-wired logic gates and accumulators in the M6802 CPU, consider one instruction. The instruction consisted of three codes (B6, 12 and 34), which occupied three memory locations in EPROM. The first byte of the instruction is a control code (B6) and the third most significant bit of this byte is logic '0', as shown. In Motorola machine code, a slightly different control code (F6) would instruct the CPU to load Accumulator B rather than Accumulator A. One could imagine that this third bit of the control code achieved the instruction by the connection shown to the router. If the control bit is '0', each bit of the 8-bit data code from memory location 1234 would be routed to Accumulator A, but when the control bit is at logic '1', these data bits would be routed to Accumulator B. Other bits in the control code likewise ensure that the overall data flow is correct.

Motorola increased the complexity of their microprocessors and introduced the 68000 chips (which were used in Macintosh computers). Other companies worked on reduced instruction set chips (RISC). The concept of RISC microprocessors was that of all the different operations that could be used in a typical program, only certain basic instructions would be used very often. Therefore, it is sensible to optimise the execution time of these basic operations rather than worry about complex chip layouts incorporating complicated operations. In this way, the overall execution speed could be increased for most program applications and the more complex operations could be accomplished by special co-processor chips.

2.3.6 INMOS transputer, T800

In the late 1980s a novel chip architecture – the transputer – caught the imagination of many UK university researchers. The transputer appeared to be the ultimate 'computer on a chip'. Its main selling point was its promise to be the solution to the speed bottleneck, which threatened conventional chip architectures. The question was 'how could one achieve the highest speed of operation or the highest throughput of data with conventional microprocessor designs like the 6802 and 68000?'. Three techniques suggested themselves. Firstly, one could increase the clock speed (which meant a continuing miniaturisation of the transistor cells in the logic circuitry). This solution is exemplified by the current INTEL processors which offer ever-increasing clock speeds. However, there has to be a fundamental limit beyond which the photolithography process used to fabricate the chips breaks down, so it cannot be a long-term solution. Secondly, one could attach more than one CPU to the address and data busses (so that one CPU can tackle one part of a problem while another CPU executes another parallel operation). Thirdly, develop special, 'application specific' chips (ASICs) which are connected to the busses and which have the capability to perform complex operations (thereby releasing the controlling CPU for other tasks).

The British firm INMOS decided to market a chip which would break the mould of this conventional thinking and provide a route to true parallel processing capable of adding more computing power as and when needed. The T800 transputer had a built-in, powerful 32-bit central processor and a floating point unit (which could perform arithmetical calculations independently of the CPU). It had on-board memory of 4kbytes and four high-speed serial links which could input or output data bytes independently of the main processor, as shown in Fig. 2.9(a). Having four high-speed links meant that the transputer could become the basic building block of complex networks, which could be configured to optimise the particular processing problem at hand. A new software language was developed for the transputer (called OCCAM) which had not only serial (SER) operations defined within it but also parallel (PAR)

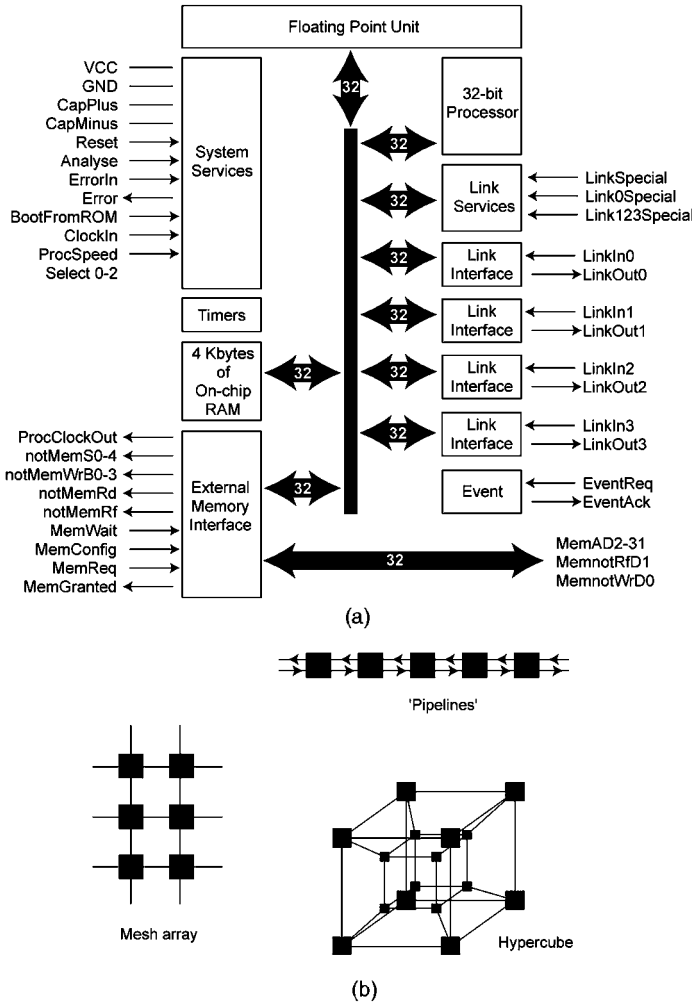


Figure 2.9 (a) A schematic diagram of the main components in the INMOS T800 transputer design is shown. Instead of an 8-bit CPU like the Motorola 6802, the transputer had a 32-bit CPU to control operations. On the same silicon chip it had timers, 4 kbytes of RAM and four high-speed serial links, which could input or output data bytes independently of the other links and any other process being performed by the CPU. The external memory interface unit allowed for gigabyte memory expansion, if necessary. (b) The four high-speed links enabled networks of transputers to be created and particular networks were more suited to solving certain types of algorithms efficiently. The mesh array was useful for finite element analysis problems. The pipelines were suited to algorithms, which could be decomposed into '*n*' operations, each requiring the same computing time to complete. The 3D hypercube network shown has the fewest communication links between any two transputers, which are the furthest apart in the network. Hence, it was suited to algorithms, which required frequent communication between any two transputers.

operations. After compilation of the software, the whole code could be tested on one single transputer (which performed the parallel operations in a time-sliced way). After testing on one transputer, the code could be modified easily to take advantage of the parallel operations by targeting a suitable network of transputers. (As described in the next chapter, the first prototype design for the Leeds 2D image analyser system contained a small network of transputers.)

Parallel processing ideas were particularly attractive for large-scale simulation and image processing applications, e.g. pipelines, mesh array and hypercube designs, as shown in Fig. 2.9(b), could be used to speed up the processing of pixel data. The pipeline design was particularly useful when the overall algorithm could be split up into roughly 'n' equal tasks (i.e. each task takes, say, ΔT secs). By connecting n processors in series and arranging for new data to be entered every ΔT secs, this pipeline would produce answers every ΔT secs (once the pipe had been filled). The mesh array design interconnected processors so that each processor simulated what was happening at a certain region of space in the real world. For example, using a finite difference approach to the solution of partial differential equations (like Poisson and Laplace equations), physical conditions could be simulated and the response of the physical system predicted. The hypercube architecture was useful if the algorithms necessary to solve a physical problem involved significant communication between all of the processing elements. In the n-dimensional hypercube design, any one processor is only 'n' communication links away from any other processor, so that the time for transfer of data is minimised in this design (see Fig. 2.9(b)).

Sadly, the transputer was gradually overtaken by the Intel series of processors. Firstly, the Pentium processor achieved ever-increasing clock speeds (whereas the transputer was never reliably manufactured for a clock speed above 25 MHz). Secondly, many researchers and most companies saw an easier upgrade path by the Pentium route (i.e. they did not have to persuade their software engineers to learn a radically new language). They could simply replace the same type of chip into existing designs. As the microprocessor clock speeds were doubling every 12 to 18 months, the perceived need for truly parallel solutions using transputers rapidly dwindled in all but the biggest machine installations and for all but the most complex problems.

2.3.7 Digital signal processor (DSP) chip

However, there were some aspects of the parallel processing ideas which did survive this era. Whereas many transputers were needed for fast parallel architectures, some American manufacturers like Texas Instruments decided to develop digital signal processor chips, which consisted of internal switching networks, multiple memory blocks and multiple dedicated processors on single chips. In effect, these were parallel architectures on a single chip. Especially in

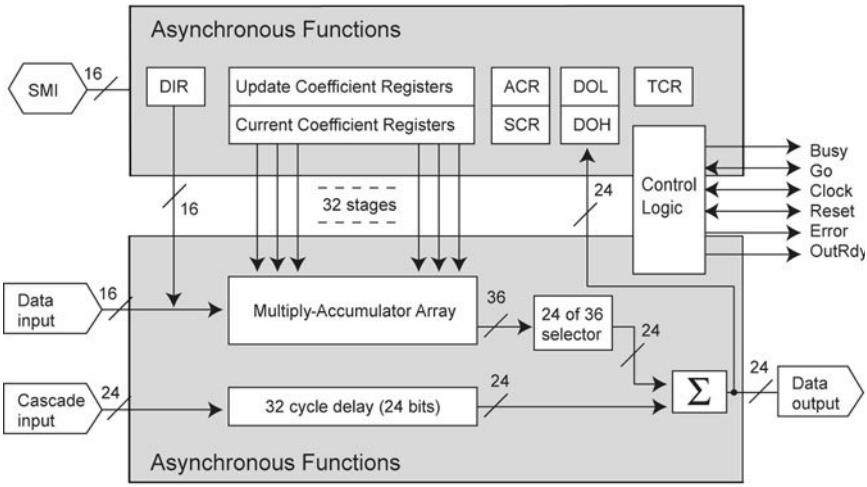


Figure 2.10 A schematic diagram of the key elements in a digital signal processor (DSP), the A100, which was also manufactured by INMOS in the 1980s.

image processing, as we shall see later in this chapter, many simple, repetitive tasks often need to be applied to each of the pixels in the digitised picture. Hence specialised chips called digital signal processors (DSPs) were developed and there is a vast literature on these chips and their applications.[8] These chips are ideal for certain, computationally intensive, image processing applications, i.e. to perform specialised, but simple, mathematical routines on pixels at videorate speeds.

INMOS also manufactured a typical DSP, called the A100, which can be thought of as a pipeline of interconnected multipliers, delays and accumulators, shown schematically in Fig. 2.10. Inside the A100 is a 32-stage pipeline of these basic cells and their importance cannot be stressed enough in a wide range of digital operations, which are often performed on image arrays. Imagine a digital signal coming in on line D_{IN} where an original analogue signal has been digitised every ΔT secs which is the same delay as in each cell of the pipeline. Therefore, at time ΔT, the output will be $O_1 = D_0 w_0$, at time $2\Delta T$, the output will be $O_2 = D_0 w_1 + D_1 w_0$ and so on where w is a weighting factor.

The ability to multiply and add in this manner enables the DSP to simulate many actual electronic circuits. For example, the output response of a simple, RC, low-pass filter (described more fully below) is shown in equation [2.6]. Because the output response of this low-pass filter depends on previous values of the input, it can be expressed in terms of a finite series, which the pipeline of 'multiply-add and delay' cells can generate.

$$V_{out}(n) = b_0 V_{in}(n) + a_1 b_0 V_{in}(n-1) + a_1^2 b_0 V_{in}(n-2) + \ldots \qquad [2.6]$$

Similarly, 2D kernels (or digital operations) for image filtering, convolution and correlation filtering and Fourier transforming (which are discussed later in this chapter) can all be undertaken by this type of pipeline structure.

Many parallel processing solutions for handling digital data require very simple but repetitive operations to be performed on data streams. In particular, the role of multiplication and additions is important for implementing partial derivatives (gradients) on digital data, and for correlation and convolution processing of data and filtering operations too.[9] As a concrete example, consider how a pipelined set of multiply–add cells could be used to perform the same function, on an incoming signal, as a simple, first order, low-pass filter. This simple low-pass filter consists of a resistor, R, and capacitor, C, as shown in circuit in Fig. 2.11(a) and the frequency response of the voltage gain is shown in Fig. 2.11(b).

Consider that the input and output voltages are sampled with a sampling interval, T, so that $V_i(nT)$ and $V_o(nT)$ represent the values of V_i and V_o at time, $t = nT$. If T is sufficiently small, the derivative can be written

$$\frac{dV_o}{dt} = \frac{(V_o(nT) - V_o(n-1)T)}{T} \qquad [2.7]$$

As the differential equation describing V_o in terms of V_i for the low-pass filter is given by

$$V_o(t) + RC\frac{dV_o(t)}{dt} = V_i(t) \qquad [2.8]$$

we can write this equation in the digital form

$$V_o(nT) + \frac{RC}{T}.V_o(nT) - \frac{RC}{T}.V_o((n-1)T) = V_i(nT) \qquad [2.9]$$

And, by rearranging,

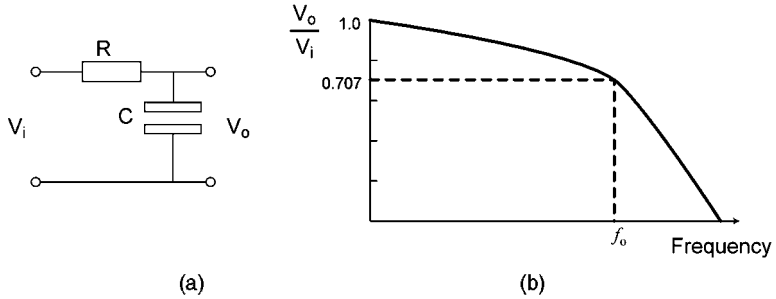


Figure 2.11 (a) A resistor (R) and capacitor (C) combination acts like a low-pass filter, which allows only low frequency signals at V_i to be outputted at V_o. (b) The voltage gain frequency response is shown and the bandwidth of this filter is given by f_0 where $f_0 = 1/2\pi RC$.

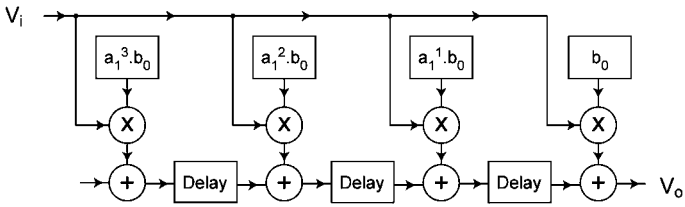


Figure 2.12 The low-pass function can be simulated digitally by a series of 'multiply–add and delay' units as shown (and see main text). These units within the A100 could easily be interconnected, as shown, to give any desired level of precision at the output.

$$V_o(nT) = \frac{1}{1+(RC/T)}.V_i(nT) + \frac{RC/T}{1+(RC/T)}.V_o((n-1)T) \qquad [2.10]$$

which is a recursion formula of the form

$$V_o(n) = b_o V_i(n) + a_1 V_o(n-1) \qquad [2.11]$$

where

$$b_o = \frac{1}{1+(RC/T)} \qquad [2.12]$$

and

$$a_1 = \frac{RC/T}{1+(RC/T)} \qquad [2.13]$$

If this equation is used for successive substitutions of the V_o terms on the right-hand side of the equation, it can be readily shown that

$$V_o(n) = b_o V_i(n) + a_1 b_o V_i(n-1) + a_1^2 b_o V_i(n-2) + a_1^3 b_o V_i(n-3) + \ldots \qquad [2.14]$$

The accuracy of the series solution depends on the number of terms used. One type of digital signal processor on the market recently was the INMOS A100 and, in Fig. 2.12, the implementation of this series expansion is shown mapped onto the 'multiply–add and delay' cells within this chip. By inspection of this circuit, it can be seen that it mimics the low-pass filter series and hence behaves like the simple analogue RC filter. Similarly, other analogue filters can be mimicked by DSPs and even filter characteristics (which are not amenable to an analogue design) can be represented in DSP form.

2.3.8 Correlation and convolution

There are other important applications of multiply–add cells which are often met in image processing problems. The 3×3 convolution kernel for an image neighbourhood can be expressed as

$$g(x,y) = \sum_{k=-1}^{1} \sum_{j=-1}^{1} h(j,k) f(x-j, y-k) \quad [2.15]$$

where the kernel coefficients are $h(j,k)$ and the pixel values are given by $f(x-j, y-k)$ around the central image pixel $f(x,y)$ and the correlation of a $n \times m$ template with an image neighbourhood around $f(x,y)$ is given by

$$g(x,y) = \sum_{k=-n/2}^{n/2} \sum_{j=-m/2}^{m/2} h(j,k) f(x+j, y+k) \quad [2.16]$$

Note the similarity between these two expressions and the fact that they use double summations of products. If a CPU within a microprocessor were to be used to carry out these multiplications and additions, significant kernel sizes could not be handled in real time. The DSPs are ideally suited to carry out these operations through their use of pipelining and dedicated hardware multiplier circuits.

2.3.9 Future computing possibilities

Over the past 30 years, microprocessor clock speeds have increased from 1 MHz to the latest Pentium running at a frequency of 2 GHz. This has been achieved through progressive miniaturisation of the basic transistor cell sizes on the silicon chips. Figure 2.13 illustrates the tremendous growth in computing power for the past 100 years.[10] However, eventually, a natural physical limit will be reached when transistors are so close together that quantum interference effects will make the motion of the elementary charge carriers unpredictable. There have been two approaches which show promise of moving to even higher performance in the medium term: the photonics route and the parallel processing route. The parallel processing options have been discussed already. The photonics route calls for the logical operations being determined by photons of light rather than electronic charge carriers. The speed of operation for current computers depends on the drift speeds of the electronic charge carriers, which constitute the current flows. These drift speeds are much lower than the speed of light. The challenge with this route therefore is to devise compact optical switching circuits but, for instance, very complex digital image operations like Fourier transforming can be achieved simply by passing the optical image through a lens (which transforms every part of the whole image in an efficient parallel manner).

In the longer term, there are two other approaches that look promising. Perhaps a fusing of biological cell technology and electronics (together with a better understanding of the structure and function of the human brain) will yield far faster computing options especially for image interpretation. Secondly, physicists are considering how to harness the quantum effect in the designs for a true quantum computer. By asking questions, they are exploring the inherently parallel

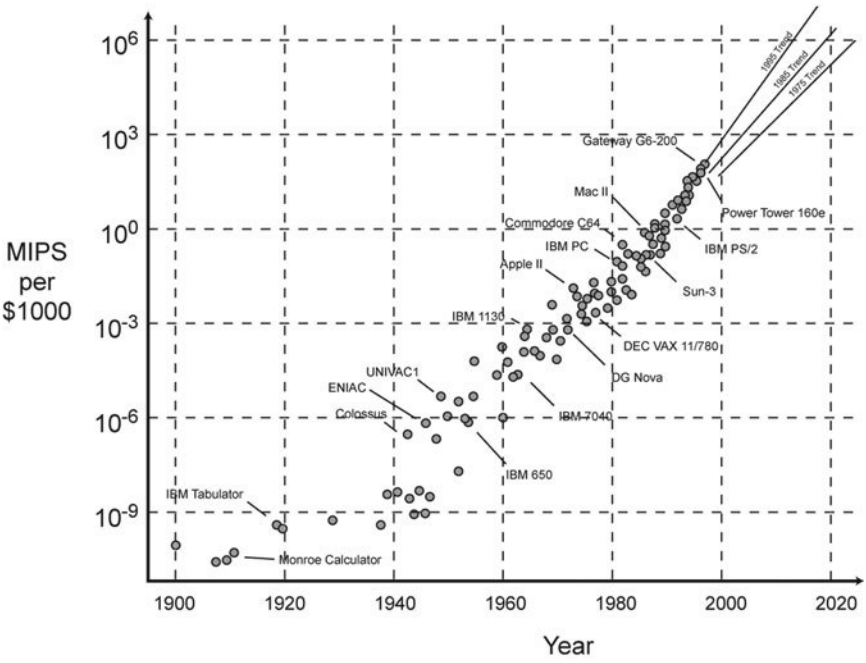


Figure 2.13 Over the past century, there has been a staggering increase in computing power (measured in millions of instructions per second) per dollar. Adapted from Moravec.[10]

nature of what we believe to be quantum states and the many possible interactions between these quantum states.[11] If a quantum computer can be engineered, a continuing reduction in size of the basic 'circuits' will actually favour higher computing throughputs rather than be a blockage to further development.

2.4 Charge coupled devices (CCDs)

CCDs are multi-channel silicon array detectors, which are designed using standard metal oxide semiconductor (MOS) architecture that is found in field effect transistors. Each pixel element is defined by three electrode gates of varied applied potential (three-phase devices). The electrodes are made from vertically stacked, conductively-doped polysilicon and overlay the photo-sensitive semiconducting silicon. The electrodes are separated from the silicon surface by a thin layer of insulating silicon dioxide. When the appropriate voltage potentials are applied to the different electrodes, the electrostatic potentials in the silicon are changed to produce zones of negative potentials surrounding a potential well. Photoelectrons generated by the incident light are then collected and stored in these localised potential wells. CCDs therefore detect and measure light in a three-step process, which is illustrated in Fig. 2.14.

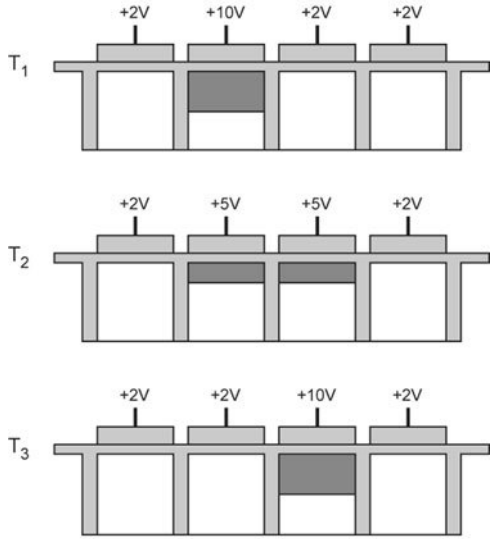


Figure 2.14 The readout operation of a charge coupled device (CCD) is illustrated. The charge trapped in one cell is transferred to the next cell by changing the voltage levels on the neighbouring cells, as shown. By repeating this cycle on adjacent cells, the charges associated with each cell (or pixel) are sequentially read out into a charge amplifier, which creates a varying output voltage signal (in effect the video line scan signal).

1. *Absorption of the incident photon energy, followed by the creation of electron-hole pairs is proportional to the number of absorbed photons*. This process is the well-known photoelectric effect. The number of electron-hole pairs generated in each discrete pixel depends on the incident light intensity. However, the number of generated electron-hole pairs is a non-linear function of the incident wavelength of the light. As shown below, the charge resulting from the generated electron-hole pairs in a given pixel is collected in the potential well of that pixel.
2. *Transfer of the resulting charge packets within the array from one pixel to the next.* At the end of an exposure, the charge integrated in each pixel is transferred from one pixel to the next in a column until it finally reaches the readout registers located along one edge of the chip. This charge transfer is accomplished by sequentially changing the applied potential to the different electrodes as shown below. During this process of parallel charge transfer, the charge accumulated in one row of pixels is simultaneously transferred to the corresponding pixels in the next row.
3. *Conversion of charge to voltage and subsequent amplification.* After the parallel charge transfer processes, the charge from each row of pixels is accumulated at the edge of the chip. The charge from the individual readout registers is then shifted to an output sense node by an analogous, serial,

charge transfer process. Here the charge is converted to an appropriate voltage. The potential level in the floating diffusion sense node varies depending on the amount of charge present at the node. The sense node is followed by a charge-to-voltage amplifier, which preserves the linear relationship between the generated charge and the voltage output.

A number of CCD architectures are now available (e.g. full frame, interline and frame transfer devices). A full frame device has its entire chip area as the active sensing area. In these scientific imaging devices, the readout registers are arranged along the bottom edge of the chip making these CCDs ideal for spectroscopic applications. A low noise signal sensing output amplifier is used to measure the accumulated charge from each register bin. The most sensitive and low noise CCDs are the ones which have Peltier coolers attached.

In order to choose the most appropriate CCD for the intended application, there are three prime considerations:

1. What is the wavelength range of interest?
2. What is the anticipated signal or light levels?
3. What is the required spectral coverage and resolution?

The answers to these three questions determine the chip format, the type of cooling and the individual pixel size. In addition to these operating parameters, there are other important experimental factors to consider for an optimal detection system. For example, the grade of chip (guaranteed maximum number of faulty pixels) is important. Also, the required dynamic range of the measurements (determined by background noise levels and the maximum charge which can be captured by each pixel potential well before saturation of the charge-to-voltage amplifier) is important. Finally, the required speed of the data acquisition (which depends upon maximum readout rates) is a vital consideration.

2.4.1 Spectral response

A good measure of the spectral response of a CCD is the quantum efficiency of the sensitive region. The quantum efficiency is defined as the percentage of incident photons on the detector converted into electrons (sometimes called photoelectrons). Examples of different chip formats are as follows:

- front-illuminated
- front-illuminated with lumigen coating
- open electrode
- back-illuminated
- front-illuminated deep-depleted

In standard front-illuminated CCDs the polysilicon electrodes are deposited on the front surface of the detector. Hence, all photons incident on the detector

must penetrate through these electrodes before reaching the photosensitive silicon substrate. As the transmittance of the polysilicon electrodes decreases below 500 nm, these front-illuminated CCDs have a spectral response between 400 nm and 1100 nm and a peak quantum efficiency at 680 nm, as shown in Fig. 2.15 – plot (a). Above 1100 nm, the photon energy is lower than the silicon band gap energy and hence they are transparent to the incident photons. These front-illuminated CCDs are easy to fabricate and hence are the cheapest. This chip format is ideal for visible light applications, e.g. photometric characterisation of visible light sources and Raman spectroscopy with visible laser excitation (at 488 nm, 514 nm and 532 nm).

If the application requires that the CCD be capable of detecting a signal in the UV part of the EM spectrum, then a thin UV fluorescent phosphor film can be deposited on the front-illuminated sensors. For example, lumigen is an organic phosphor which, when illuminated by light at 200 nm wavelength, emits photons at 550 nm (that can be detected by the underlying sensor). Fortunately, lumigen is transparent to the visible region and, hence, the spectral response is as shown in Fig. 2.15 – plot (b).

More expensive CCDs, which exhibit higher quantum efficiencies in the UV, are the open electrode CCDs. In this variation, a portion of the UV absorbing polysilicon is etched away so that a large portion of the underlying photosensitive silicon is exposed to the incident light. This technique allows the incident photons to reach the detection region unabsorbed. The typical spectral response of an open electrode CCD is shown in Fig. 2.15 – plot (c). Note that there is also an increase in near infrared sensitivity. A minor disadvantage is that these devices have reduced charge-handling capacity, but the dynamic range is not usually affected (because it tends to be determined by the analogue-to-digital converter used in the subsequent signal processing). The open electrode CCD is suited to, say, Raman spectroscopy where the excitation laser is in the red (632 nm) or near infrared (785 nm) to avoid fluorescence interference.

Back-illuminated CCDs have dramatically increased quantum efficiencies because the incident photons bypass the polysilicon gate structure. However, since the UV photons are absorbed very close to the silicon surface, the resulting generated photoelectron-hole pairs can recombine before reaching the potential wells. To minimise this effect, the back-illuminated devices are thinned to approximately 10–15 microns thick. The quantum efficiency can approach 90 per cent with the addition of anti-reflection coatings, which are optimised for different regions of the spectrum. The response shown in Fig. 2.15 – plot (d) has a coating optimised for the NIR. The back thinning process is difficult and the yields are low – hence these CCDs are expensive.

The NIR response of CCD detectors can be enhanced by modifying the photosensitive silicon substrate. In these devices, known as deep-depleted devices, the silicon substrate is made thicker to extend the sensing depth, which increases the possibility of NIR photons being absorbed well below the front

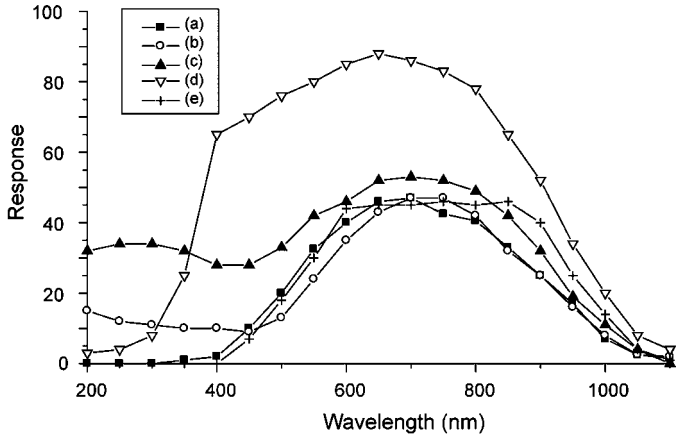


Figure 2.15 The spectral responses of different types of CCD on the market are shown: (a) front-illuminated CCD, (b) front-illuminated, lumigen coated CCD, (c) open electrode CCD, (d) near infrared anti-reflection coated CCD and (e) front-illuminated, deep-depleted configuration CCD.

surface of the silicon. To minimise the recombination of the photoelectron pairs within this thick substrate, a portion of the substrate is lightly doped with a p-type dopant (see the response in Fig. 2.15 – plot (e)).

2.4.2 Operating temperature

In terms of decreasing cost, the cooling mechanism used with CCDs is either liquid nitrogen cooled, four-stage thermoelectric cooled or two-stage thermoelectric cooled. The flux of photons incident upon a CCD during a measurement will determine the exposure time for obtaining the necessary signal above the background noise. This exposure time is ultimately affected by the dark current leakage. The leakage is expressed in either nanoamps per cm^2 or counts per pixel per unit time. The dark current is caused by thermally generated electrons and hence they will be minimised by reducing the temperature of the sensing material, i.e. by cooling the CCD. A graph of dark noise versus temperature is shown in Fig. 2.16. The two-stage Peltier (thermoelectrically) cooled devices get down to -25ºC and achieve dark currents around two to five electrons per pixel per second. This performance is usually adequate except for the most demanding applications involving detection of the lowest photon fluxes and hence long exposure measurements.

2.4.3 Spectral and spatial resolutions

In spectroscopic applications, there are two types of coverage: the spectral and the spatial coverage. The light signal to be analysed enters the spectrograph

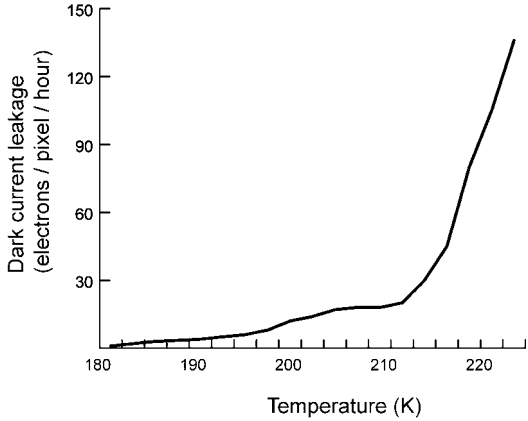


Figure 2.16 The performance of CCDs may be improved by cooling them to below room temperature, usually with miniature thermoelectric cooler units. The dynamic range is extended because the noise level (due to the dark current leakage) decreases dramatically as the temperature is reduced.

through an entrance slit and is separated into its component wavelengths by some form of dispersive element (e.g. a prism or diffraction grating). At the exit focal plane of the spectrograph, this wavelength-dispersed information is normal to the height of the entrance slit. The total spectral coverage is therefore defined as the amount of dispersed spectral information that can be collected by the CCD in a single exposure, i.e. $\Delta\lambda = \text{dispersion} \times \text{width of CCD}$.

The CCD detector captures the energy incident along the entrance slit height. When the entrance slit is fully illuminated, a taller detector (a CCD with larger height of its active area) will be able to capture more of the incident photons. Therefore, different light sources could be positioned at different heights across the CCD, e.g. several fibre optic cables could be used to collect the light from different sources. (In an astronomical context, this can be most useful because simultaneous monitoring of possible variable sources, and nearby reference sources, can effectively remove the fluctuations in background seeing.) In applications where the signal is tightly focussed onto a spot at the exit plane, it is better to use small area CCD chips in order to minimise the dark noise from the non-illuminated pixels.

Clearly the spectral resolution is given by $\delta\lambda = \Delta\lambda/N$ where N is the number of pixels across the CCD but, as three pixels are required to define a peak within background, the effective resolution will be three times this limit ($3\delta\lambda$). Also, for spectroscopy, the entrance slit (10 microns or wider) ultimately defines the amount of light entering the spectrometer and therefore having the individual pixel size smaller than the entrance slit will gain no advantage in terms of spectral resolution. In low level light applications, some CCDs allow the electronic binning together of groups of pixels to form super-pixels (which increases the dynamic range and speed of these detectors).

The chip parameters that determine spatial resolution for normal images (and spectral resolution when used in spectrographs) are hence the detector width, height and individual pixel size. Typical CCD sizes might be

- 512×512 pixel arrays with 24 micron square pixels which implies an active chip area of $12.3\,\text{mm} \times 12.3\,\text{mm}$.
- 2048×512 pixel arrays with 13.5 micron square pixels (implying an active chip area of $27.6\,\text{mm} \times 6.7\,\text{mm}$).

All CCD chips have cosmetic blemishes or defects, which are regions either with reduced sensitivity or with increased dark current. These blemishes might be individual pixels, clusters of pixels or whole columns in the array. Chips are graded according to how many blemishes there are – the most expensive chips have the fewest blemishes.

2.4.4 Dynamic range and gain

The dynamic range is defined as the ratio of the largest useful signal that can be measured to the smallest detectable useful signal. For CCDs the dynamic range is therefore the ratio of the pixel, full-well capacity to the readout noise of the sensor. (In spectroscopic work, when several pixels in a column or the entire column are binned, the maximum signal intensity is limited by the charge storage capacity of the readout registers. Hence, the dynamic range will be greater.)

Larger pixels have greater charge storing capability – a $26\,\mu m$ square pixel can store 500,000 electrons whereas a $15\,\mu m$ square pixel only stores up to 90,000 electrons. However, note that the resulting image (or 2D pixel array data) from the CCD camera has a dynamic range defined with respect to the analogue-to-digital converter (ADC) of the controller or framegrabber board within the host computer. The best ADCs tend to have 16-bit resolution, i.e. the signal can be digitised to 65,535 counts of intensity resolution. The more standard ADCs for video work have only 8-bit resolution (meaning that only 255 counts of resolution are possible). These necessary peripheral devices are going to be the limiting factor in practice.

2.5 Digitisation and ADCs

An overview of the operation of a typical ADC for digitising analogue signals from a camera is shown in Fig. 2.17. As well as the input analogue signal and the set of digital output lines (eight lines for an 8-bit ADC), there are a few important control lines to consider too. The Command Convert (CMD CNVT) input control line expects to see a pulse or change of state and this line tells the ADC when to digitise the incoming analogue signal. There are essentially two things that the user can arrange here – either provide a fixed clock signal which will make the

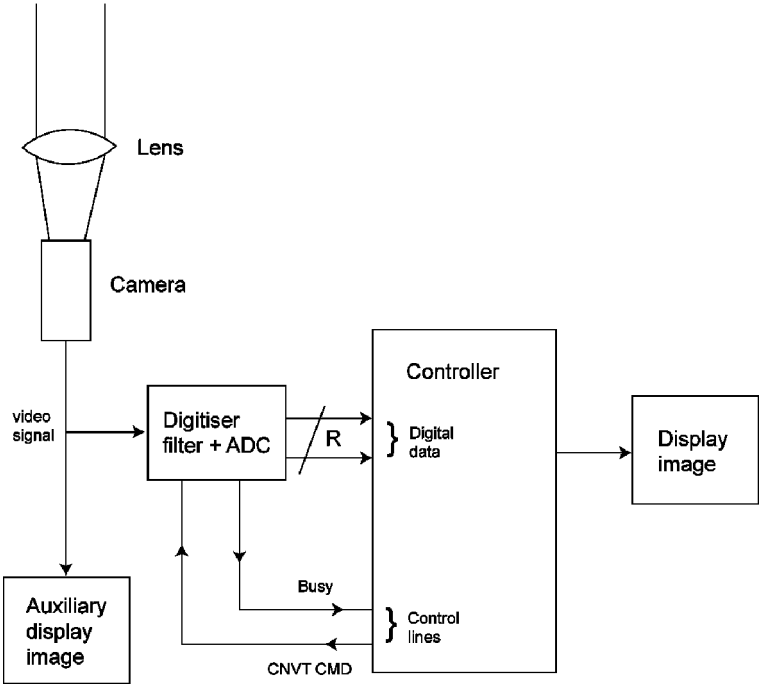
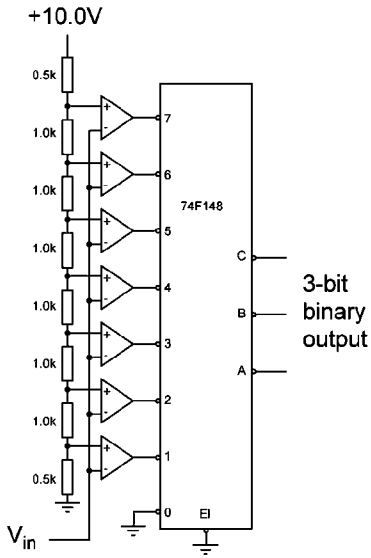


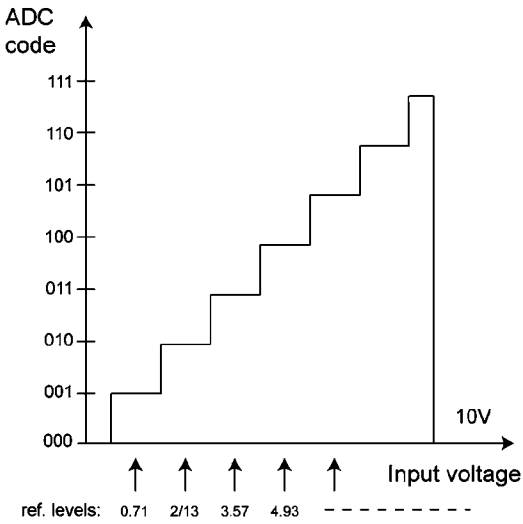
Figure 2.17 A schematic diagram illustrates the connections between an analogue-to-digital converter (ADC) and a host processor. Firstly, a command convert pulse (CMD CNVT) is sent to the ADC by the host processor. The ADC alters the state of the other control line (Busy), say to a logic 1 state, and begins to digitise the input analogue voltage into an R-bit binary code. As soon as the R output lines are stable, the ADC alters the state of Busy to logic 0. This change of state is picked up by the host processor and the R-bit pattern is transferred into memory. When the host processor is ready, it sends another pulse to CMD CNVT and the cycle repeats until the required number of digital values has been collected and stored. This process is called handshaking and it ensures synchronisation of communication between a slave device (ADC) and a host controller, even when they are operating at different speeds.

ADC digitise regularly or enable/disable a set of pulses. A Busy output control line is also provided whose state informs the computer or memory device waiting to receive the digitised data that the digital output lines are now stable and can be read. For example, this line may go to logic high state when the ADC is digitising and the receiving device will wait for the control line to go to logic low state before it reads the 8-bit output lines. The flash ADC design has proved to be effective within video sampling devices and a simplified view of its functional form may be discussed with reference to a 3-bit design, as shown in Fig. 2.18.

Imagine the incoming signal being compared to a set of equally spaced, reference voltages. These reference voltages are at the non-inverting inputs of the seven op-amp comparators. If the signal voltage is lower than the reference



(a)



(b)

Figure 2.18 In order to digitise the analogue signals from a video camera sufficiently quickly to create a 512×512 digital image on screen, which follows video rates of 32 frames per second, the ADC has to digitise within 100 nanoseconds. (a) The outline of a fast 3-bit ADC is shown – called a flash-ADC – where the input voltage is compared to the voltage levels in a potential divider network. Depending on the voltage, some of the inputs to the encoder are set to logic 1 and the remaining inputs stay at logic 0. (b) The encoder outputs a 3-bit code depending on how many encoder inputs change state. The ADC output code and the input signal voltages, which trigger these codes, are shown.

voltage, the output of the comparator will be in a logic high state. In the short time taken for the input comparators to switch states, a binary code will appear at the input to the encoder whose value is linearly related to the incoming signal voltage level. The encoder (called a priority encoder) acts like a look-up table, transforming one of the eight different voltage levels to a unique 3-bit code on the output lines A, B and C after a few logic gate switching times. Consider that the input signal had a value of 2.5 volts. The reference voltages at the non-inverting inputs of the bottom three comparators are $0.5 \times 10/7 = 0.71\,V, 1.5 \times 10/7 = 2.13$ V and $2.5 \times 10/7 = 3.55$ V. Hence, the 8-bit code at the input to the encoder will be 11111000 and the output from the encoder will be 011 (to represent that the lowest three levels only were exceeded).

Instead of a small reference voltage network, which contains only these eight voltage levels, an 8-bit flash ADC chip will have a set of 256 voltage levels to compare to the incoming signal voltage. However, the time for conversion of the incoming signal into an appropriate 8-bit output will be virtually the same as for this simpler model of the flash ADC.

2.5.1 Slow scan versus real time

If more economical circuit designs are required or if, say, 16-bit pixel intensities are necessary for long time exposure studies in astrophysics research, there is an interesting alternative to true video rate sampling of the signals from the CCD sensor. A slow scan approach can be adopted in which one sample of the pixel intensity is made every line. As shown in Fig. 2.19(a), the digitisation process, and subsequent sending of byte(s) to memory, could take around 40–50 microseconds. However, there will still be time before the next digital sample has to be taken at the same x position on the next line scan. Sometime later, the pixel columns from $n = 1$ through to $n = m$ have been sampled and displayed, as shown in Fig. 2.19(b). Hence, although the process takes a few seconds to recreate the complete 2D scene at the highest 16-bit resolutions (hence the term slow scan), this may not be a problem for the researcher. For example, if the application requires a long time spectrographic or deep-field exposure on an astronomical object, it does not matter to the astronomer if the readout takes seconds to complete. Such higher resolution studies enable the user to take full advantage of the wide dynamic range and low noise capabilities of the CCD.

2.5.2 Quantisation error in ADCs

An ADC is a device which accepts an analogue input x_a and produces a digital number, x_d (usually as M parallel binary signals). The range of x_d is 0 to $(2^M - 1)$ in uniform steps where the analogue equivalent to each step is a quantising interval q. The largest full-scale analogue signal is therefore $(2^M - 1)q$ and the smallest change that can be reliably observed is $\pm q$. The

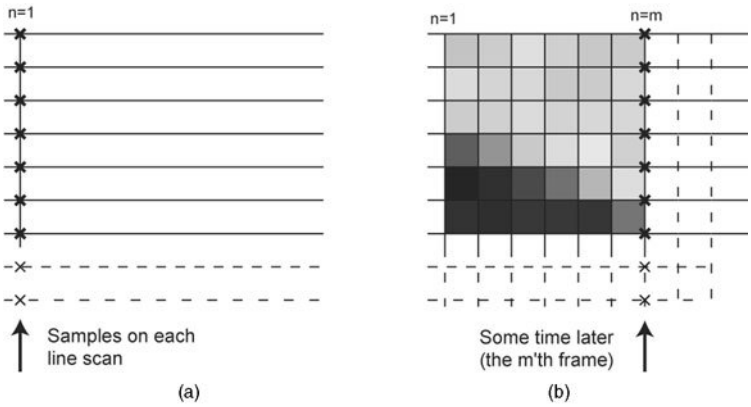


Figure 2.19 In some research areas, it is necessary to digitise the intensities of wide dynamic range optical devices, like CCDs, to 16-bit precision. ADCs, which are capable of this precision, generally take many microseconds to perform the digitisation. One technique for achieving image creation is the slow scan technique. (a) A sample is taken every τ microseconds, where $\tau = 53$ μsecs is the time taken for a line scan, as shown in Fig. 2.21. (b) Hence, the image is gradually built up, column by column, within a few seconds.

relationship between x_a and x_d is shown in Fig. 2.20(a), along with the error between output and input, $e = x_d - x_a$. It can be seen that the error lies between 0 and $-q$. If the input is offset by $q/2$, as shown in Fig. 2.20(b), the error becomes alternatively positive and negative and it is in the range $-q/2 < e < q/2$. It can normally be assumed that the error is a random quantity which is distributed uniformly over the interval $0 < e < q$ or $-q/2 < e < q/2$, i.e. the probability of error is $1/q$ over the digitisation interval, q.

It can be shown that the mean error is either $-q/2$ or 0 for the two cases shown in Fig. 2.20, but the variance, σ^2, for both cases is given by

$$\sigma^2 = E\{(e-\overline{e})^2\} = \int_{-\infty}^{\infty} (e-\overline{e})^2 p(x)dx = \int_{-\infty}^{\infty} (e-\overline{e})^2 . \frac{1}{q} . de$$

$$= q^2/12 \qquad [2.17]$$

where $E\{\ \}$ denotes the expectation value of the mean square error function. Hence, even if the incoming signal is noise free, the digitisation process introduces uncertainty and hence an effective digitisation error.

2.5.3 Timing implications

Most cameras have a standard video output signal or have the ability to output pixel values as digital bytes. Assume that the sensor in the camera produces a

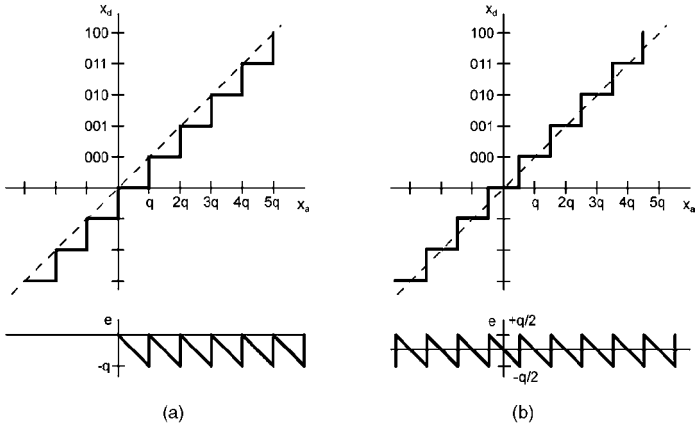


Figure 2.20 (a) The relationship between input voltage, x_a, and output digital code, x_d, is a 'staircase' function. If the digitisation steps are of size q, it follows that the digitisation process will introduce the error function shown, as the input voltage is increased. (b) By introducing an offset of $q/2$, the error function oscillates about an error value of zero and hence the mean error will be zero. However, the variance is the same for both cases, showing that any digitisation process will introduce an error of this kind.

RS170, video signal waveform.[12] Figure 2.21 illustrates a typical first line of the video output, after the frame synchronisation pulse. The front porch, horizontal sync and back porch is shown together with the 53 microsecond analogue signal, which represents the variations in intensity over a horizontal line of a 2D image. Therefore, if there are to be 512 picture elements (pixels) or samples along such a line, the sampling rate (together with saving of the digitised value in memory)

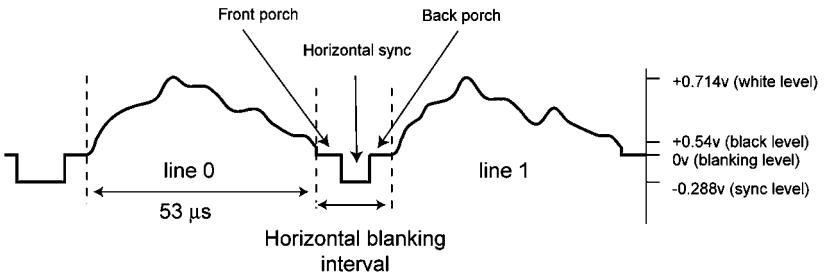


Figure 2.21 A typical schematic of the first two line scan signals from a video camera. The start of a new image frame is synchronised by a negative-going, longer duration, frame sync pulse. After a short period, the variation in analogue signal from 0 volts to +0.714 volts corresponds to the light intensity variations along the first line of the image. After another short delay, a negative-going, horizontal sync pulse synchronises the start of the next line and the process repeats until the whole image frame has been sent.

must be approximately every 100 nsecs (i.e. 53/512 μsec). This sampling rate will allow real-time video to be sustained and images to be updated 32 frames every second.

2.5.4 Nyquist sampling

An essential idea associated with digital sampling is that the information content within the original continuously time varying signal cannot be fully retrieved. The lower the rate of sampling, the more uncertainty must be introduced when recreating the original signal. This effect can be appreciated when considering a signal, which contains a dominant sine wave at a frequency, f Hz and hence a period, $T = 1/f$ secs, as shown in Fig. 2.22(a). Clearly, if f samples per second are recorded, when it comes to replaying the digitised values, the sine wave will not be seen – only a constant voltage value (somewhere between $\pm A$, where A is the amplitude of the original waveform)! If the sampling rate is between f and $f/2$ samples per second, the replayed signal will display an artefact, which is a

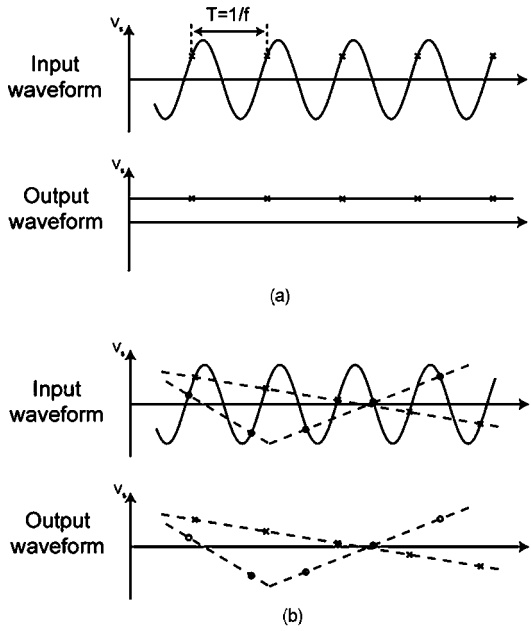


Figure 2.22 The important principle known as the Nyquist sampling theorem is illustrated. (a) The original sine wave signal, with period T is shown together with the resultant non-varying, output waveform created when a sample is taken of the input signal once per period. (b) Two more output waveforms are shown, for an intersampling time greater than the period T. Note that, in each of these cases, the output waveform bears no resemblance to the original input signal!

repetitive waveform at an incorrect frequency, as shown in Fig. 2.22(b). The Nyquist sampling theorem states that there must be at least $2f$ samples per second, if the highest signal frequency is f Hz, in order to recreate the original signal. In order to prevent these artefacts, most image processing systems have anti-aliasing filters in the incoming signal line to ensure that all frequencies greater than $f_s/2$ Hz are removed (where f_s is the digital sampling rate). Therefore, if video rate sampling is adopted, low-pass anti-aliasing filters having a cut-off frequency of 5 MHz will be used at the inputs to front-end framegrabbers (as discussed later).

2.5.5 C80 and the Genesis framegrabber

A schematic overview of Texas Instruments' C80 digital signal processor is shown in Fig. 2.23. This chip incorporates a mix of parallel processing features, which allow multiple instructions to be implemented and multiple datasets to be routed and processed at the same time. The different constituents of this chip are:

- a RISC master processor
- a cross-bar switching network
- four separate memory banks
- four parallel processor – advanced digital signal processor – units
- a transfer controller for rapid access to external memory elements
- two video controllers.

The cross-bar network allows any of the four processors to be connected to any of the memory blocks (in order to process information which is held there).

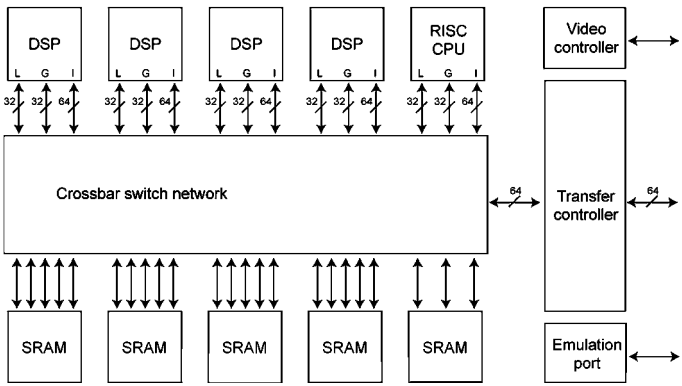


Figure 2.23 A schematic overview of Texas Instruments' multiple, digital signal processor (DSP) chip, the C80. This chip has its own reduced instruction set (RISC) processor on board to control operations and data flow between the DSP units and the bank of SRAMs. The cross-bar switching network enables any DSP to be connected efficiently to any memory block.

Unlike the transputer concept, where parallelism was achieved via the high-speed links between different transputers, there is significant parallelism on the chip itself.

The advanced digital signal processors on the C80 have special features, e.g. long instruction words (64 bits) allow up to 15 RISC-equivalent operations to be specified in a single instruction. It has a 16×16 multiplier, a 32-bit ALU (arithmetic logic unit), dedicated adders for either address generation or for arithmetic operations and single clock cycle access to on-chip memory. This last facility allows two 32-bit data transfers per processor in every cycle concurrent with data operations. These features enable the chip to execute two million RISC-like operations per second and it can sustain a 2.4 gigabyte/second data transfer rate.

Each of the four memory blocks contains 8 kbytes of data and an extra 18 kbytes of RAM is on chip too. Because each memory access takes only one cycle, inter-processor communication protocols such as pipelining can be implemented easily in software. If any two processors try to access the same memory location at any time, the cross-bar connections ensure that such contentions are resolved through a priority-based scheduling scheme.

The two video controllers generate video timing signals and also allow VRAM transfer cycles. Each controller contains a frame timer and associated circuitry, which eliminates the need for external circuitry (reducing the board space and number of chips to implement a system). The clock speed for the C80 is 50 MHz.

One of the latest, high performance, image framegrabbers, the Matrox 'Genesis', as shown in Fig. 2.24, incorporates this powerful C80 chip. Matrox have also designed a neighbourhood operations accelerator (NOA) chip to offer the highest speed image processing functions (which are discussed later in this chapter). This board interfaces to virtually any camera or input device. It can capture up to four video streams from standard monochrome (RS-170/CCIR) cameras or video from one RGB camera. It has a digital interface enabling TTL/RS-422 synchronisation and control and it can access slow scan devices too. Anti-aliasing low-pass filters are used and also ADCs which can sample up to 140 MHz. Look-up tables after the ADCs in each channel manipulate the analogue or digital video data before being passed on to the main board for image processing.

A video interface chip (VIA) on the main board even allows the acquisition of digital RGB colour line scan cameras that provide colour component data serially (one component per clock cycle) rather than in parallel. This Matrox Genesis framegrabber can achieve 100 billion (10^{11}) operations per second. The Genesis can have its range of functions extended by the addition of other, dual C80-based processor boards, which are interconnected by a high-speed, 32-bit wide Peripheral Component Interconnect (PCI) bus.

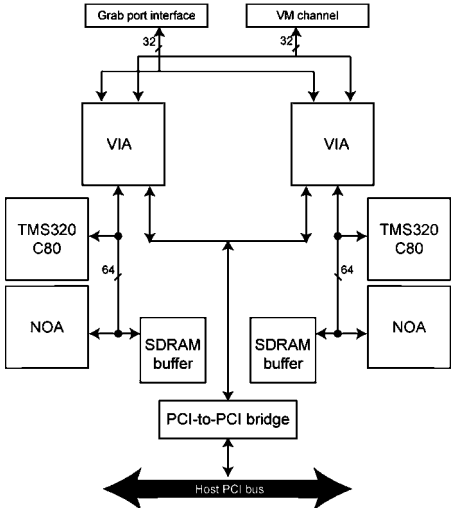


Figure 2.24 Overview of some of the functions of a modern framegrabber board for digitising video camera signals, called the Matrox 'Genesis' framegrabber. In this design, two digital signal processors (C80s) are used and two special neighbourhood operation accelerators (NOAs) to speed up the processing of images with $n \times n$ convolution filters (discussed later). Images are stored in SDRAM memory units and fast parallel transfer of the image bytes is achieved via the high-speed PCI bus.

2.6 Digital images

Interest in digital imaging dates back to well before the arrival of digital computers. In the early 1920s a submarine cable was used to transmit newspaper pictures between London and New York. This system employed some very basic ideas of converting image data into numbers but the picture quality was extremely crude and the transmission took three hours. However, this was a vast improvement on the manual image delivery method which could take a week or more. Today, the field of digital imaging covers a wide spectrum of applications within medicine, physics, astronomy, biology and just about any applications which involve images. Furthermore, digital imaging has now found its way into the household with many camera manufacturers selling affordable digital cameras.

The term image processing is applied to a broad range of applications and processes, with subsets of image processing including machine vision and machine perception. It is probably best defined as any modification or interpretation of a digital image by a computer. The image being processed does not have to be one acquired by a digital imaging device, however this is quite often the case. Hence, the terms digital imaging and image processing are closely related.

Image processing can be a complex task both computationally and mathematically. For this reason many researchers think that it is best avoided.

However, in the particular research area we are interested in – the microscopy of materials – the images observed are frequently of high contrast with regular, well-defined features. Fortunately, these attributes are desirable because they greatly simplify the task of processing such images. So take heart, all is not lost! The previous sections described the acquisition of images by a CCD array and their transfer to the computer memory. From here on, the image is in its digital form and all subsequent processing is performed by the computer.

2.6.1 The structure of digital images

In microscopy, a continuous two-dimensional image $f(x, y, B)$ is formed where f is the colour with brightness B at some point (x, y). A continuous function can be stored digitally by the process of sampling as illustrated in Fig. 2.25. Figure 2.26 illustrates the representation of an image in digital form. A monochrome digital image can be thought of as a matrix with the row and column indices representing a point in the image and the corresponding matrix element identifying the brightness at that point. The individual elements of this digital array are termed picture elements, which is usually abbreviated to pixels. In Fig. 2.26 the top-left position of the image origin and the xy axis convention is illustrated. The direction of the y-axis may seem counter-intuitive, however this convention is used in the majority of imaging software (although there are always some exceptions). The perceived quality of a digital image is determined by the size of the image matrix and the quantisation of brightness levels.

The number of pixels used to represent an image define its resolution. A typical digital image may have dimensions of 800×600 resulting in 480,000 individual pixels, although other dimensions/resolutions are possible. Most digital image acquisition devices capture images with an aspect ratio of 4:3, therefore digital images are often found to have this aspect ratio. This choice of aspect ratio dates back to the 1930s when it was set as a standard for early motion pictures by the Academy of Motion Picture Arts and Sciences. This

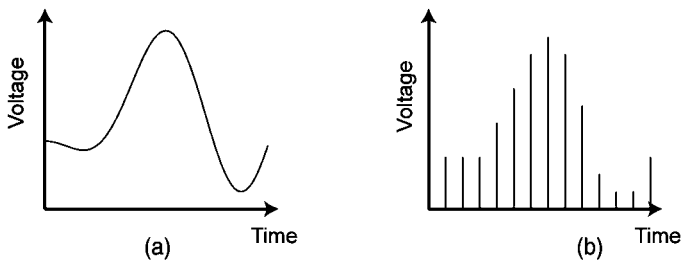


Figure 2.25 (a) A continuous analogue signal voltage, which may, for example, correspond to time-varying brightness on a light sensor. (b) The digitisation process, where the input signal has been digitised by an ADC, i.e. it has been quantised in both magnitude and time.

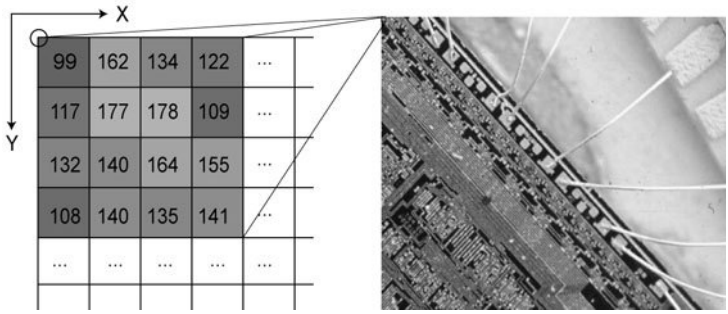


Figure 2.26 The storage of a digital image. In an 8-bit greyscale image, the intensity of each square picture element (or pixel) is stored as a value in the range 0–255, where 0 denotes a black pixel and 255 denotes a white pixel.

clearly influenced the design of televisions and the standard aspect ratio was chosen in the 1950s by the National Television Standards Committee. Computer monitors are based on televisions and it is no surprise that the 4:3 standard was continued. In order that screen pixels are square, monitor resolutions also have this same aspect ratio, hence digital images are often captured and displayed with this aspect ratio.

The number of brightness steps used to store each pixel's intensity is usually termed the colour-depth and is quoted in bits. For example, each pixel in an 8-bit image has a possible 2^8 ($=256$) different grey levels and the range of values possible is termed the dynamic range of the image. The standard unit of computer storage is the byte, which contains 8 bits of information, hence 8-bit images may be conveniently stored with one byte per pixel (or three bytes per pixel in full-colour RGB mode). The human eye can distinguish somewhere between 20–30 brightness levels in a monochrome image, therefore an 8-bit colour depth is more than adequate for capturing images for visual presentation. The single byte storage and adequate dynamic range of 8-bit images have made them the most widely used image format for a number of years. However, further processing of an image may alter the image dynamic range. For example, histogram manipulation and spatial filters affect the dynamic range globally and locally, respectively. Therefore, brightness quantisation may become visible in the processed image and, hence, some modern framegrabbers allow capture of images with higher colour depths.

The choice of a suitable image resolution for digital image storage is vital. The reasons for this become self-explanatory when some of the problems associated with under-sampling are studied. The field of digital signal processing pre-dates digital imaging and the problem of storing a continuous (analogue) signal in digital form are common to both. The Nyquist sampling theorem was introduced previously in the context of sampling an analogue signal. The problem of aliasing is present in all forms of sampled data, for

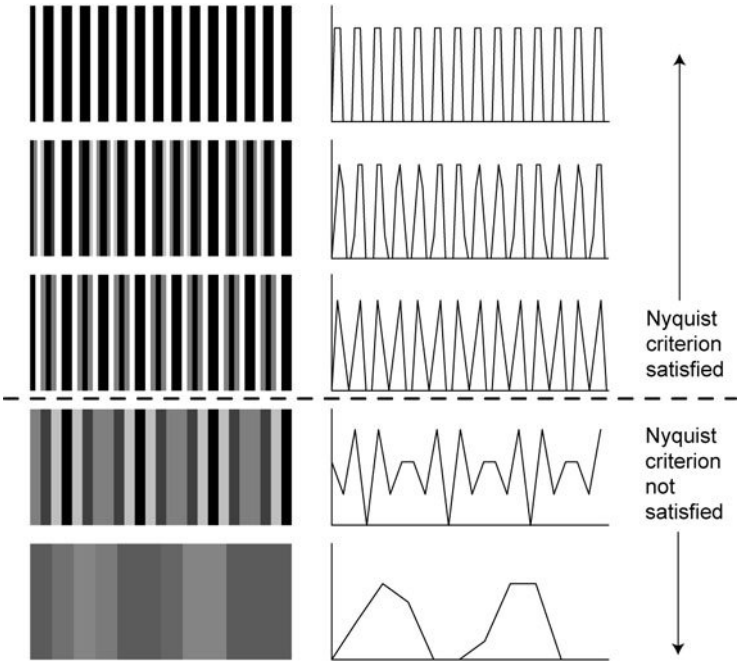


Figure 2.27 An image of a grating displayed at a number of different resolutions (i.e. different digital sampling rates). When the spatial resolution, as defined by the sampling rate, is lower than the Nyquist criterion, the apparent spatial frequency of the grating's spacing appears lower – important spatial information has been lost.

example, in video temporal aliasing may occur, where the wheel of a car seems to turn backwards. The Nyquist criterion states that a signal must be sampled at least twice as fast as the bandwidth of the signal to reconstruct the waveform accurately; otherwise, the high-frequency content will alias at a frequency inside the spectrum of interest. Aliasing is the appearance of false, lower frequency information when data are sampled. In digital imaging, the most frequently observed form of aliasing is spatial and an example is shown in Fig. 2.27. The result of imaging a regular, black and white grating pattern is illustrated at a number of different spatial resolutions. In digital images the Nyquist criterion can be applied as follows, the spatial resolution of the digital image must be at least twice that of the features that we wish to resolve within the image. Therefore, with the grating example shown in Fig. 2.27, the image resolution, i.e. the pixel widths, must be half (or less than half) the grating spacing.

In practice, there are a number of factors which limit the maximum achievable resolution. The amount of storage space, both long-term (on hard disk) and short-term (in RAM), that an image requires is resolution dependent. However, advances in computer technology have made memory a cheap commodity and hence storage requirements rarely limit an image processing

system. The most significant limiting factors on image resolution (and colour depth) are the CCD and ADC (framegrabber) that have been chosen.

2.6.2 Image processing

The practical application of quantitative measurement by image processing has many parts. By means of introduction, a particular image processing route will be considered, as illustrated in Fig. 2.28. This example focusses on a binary image processing route where the sample of interest contains discrete objects of interest. Other processing schemes are possible, but this scheme is commonly used and is therefore appropriate for illustrative purposes.

Initially the microscope is set up to inspect the area of interest within the sample. It is of primary importance that this is done carefully in order to deliver the best possible image to the computer. Any shortcuts taken to image more

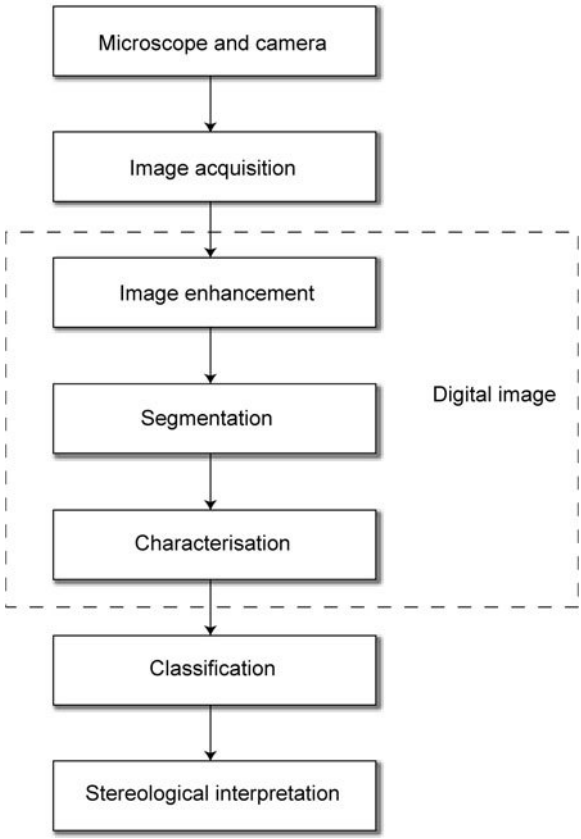


Figure 2.28 An overview of the processes involved with the manipulation of data in a typical image processing application.

rapidly will certainly result in problems further down the line, ultimately resulting in less accurate data.

Image acquisition is the process of taking an image produced by the microscope and turning it into a digital image. The digital image differs from the images observed through the eyepiece of a microscope in that they are not continuous. Digital images are composed of a number of small cells named picture elements, which abbreviates to pixels. These images may then be stored and manipulated by a computer.

Image manipulation encompasses a range of transformations, which may be applied to the digital image. These functions affect the individual pixel values of the image, the aim being to make it easier to perform the subsequent analysis on the image. For example, an image may be manipulated in some way to increase the visibility of a certain feature of interest or conversely to suppress an undesirable feature, such as image noise. It is also possible to manipulate an image in order to correct certain defects introduced by the microscope optics or image capture device, effectively enhancing the performance of the system.

Segmentation involves the identification of features of interest within the image, boundaries are found between the regions and objects within the image. As human beings we are adept at performing segmentation in real time, however computers are not nearly as well equipped. With binary image processing, this step simply involves the application of an intensity threshold. Bright regions are assigned a value of '1', which means they are of interest, and a '0' is assigned to those below the threshold which are not of interest to the user. However, the success of the segmentation process, which uses such a simple algorithm, relies heavily on the effectiveness of the prior image acquisition and image transformation processes.

The characterisation process takes the objects that have been identified by segmentation and makes a number of measurements on each. The characteristics of interest depend on the application being developed and can include length measurement, ellipticity and form factor to name but a few. After the characterisation process, the digital image is no longer required and, hopefully, all of the required information has been extracted from the image at this stage.

In applications where a number of different objects are being inspected within the same image, classification is required in order to place each object into its respective class. The characteristics of each object recorded in the previous step should give enough information about each object for this step to be possible. Even if the images being processed contain only a single object type, classification may be employed to remove rogue features introduced during the sample preparation or image processing phase.

Finally, the principles of stereology are applied in order to extract statistically unbiased conclusions from the data. The effects of our chosen image processing route on the measured characteristics must be considered together with the way in which the images are being acquired.

2.6.3 Implementation of image processing

After a sequence of image processing algorithms has been chosen, it must be implemented on a computer. There are a number of decisions that must be made:

- whether to adopt a serial or parallel computer architecture?
- which operating system to use? And, possibly, the trickiest decision to make,
- which programming language to use?

Fortunately, the financial budget available to the research project can help eliminate some of the possible answers. Furthermore, note that the outcome depends on the image processing algorithms employed so the results should be the same regardless of the implementation.

In Section 2.3 the evolution of the modern computer and the execution of machine code by the central processing unit (CPU) were described. However, direct programming of the CPU in machine code (or assembly language, its symbolic representation) is a time-consuming and laborious task. The number of individual commands available to the programmer varies from processor to processor but is typically very few, usually less than 100. Therefore, simple mathematical operations such as multiplication and division are composed of many assembly language commands.

The unfriendly nature of low-level programming was quickly addressed, and now there is an overwhelming choice of more friendly programming languages. These high-level languages have much larger instruction sets, which are often in English. However, the CPU does not understand these new languages and they must be converted to its native machine code. This process can be performed in one of two ways: compilation, where the entire program is translated into machine code in one go, or interpretation where individual lines or commands are translated as they are executed. Clearly, the speed of execution is dependent on the efficiency of the compiler or interpreter respectively.

One of the first steps in developing an image processing routine is actually getting at the data within the image. Most commercial framegrabbers are provided with a library of routines that interface with a range of programming languages. This library will probably contain a simple command, which captures the current image and places it into a 2D array in the computer's RAM.

2.7 Storage and retrieval of images

Before the advent of digital imaging, the storage of images acquired using a light microscope was a long-winded process involving the exposure and development of photographic plates or film. Before the invention of camera film, image storage was an even more laborious task requiring patience and a microscopist with more than just a little artistic ability. The line drawings of pre-nineteenth-

century researchers are sufficiently illustrative but quantitative measurements from these drawings would certainly involve a high degree of error.

During the last decade the introduction of digital capture devices has revolutionised the way images are captured and stored with all kinds of optical devices including microscopes. Digital imaging methods enable almost instant recording of images and (providing the required file can be found) similar retrieval times. With this increased rate of image capture, it is not atypical for a microscopist to record hundreds of images digitally during a single working day. However, this ease of capture can cause complications. There are many different graphical formats for storing image data and the wrong choice can render an image virtually useless. Furthermore, someone wishing to manipulate the information held within the image needs to know the storage structure of that particular format in order to interpret the data.

2.7.1 The structure of image data files

The human eye images the world in a similar fashion to a CCD camera, with the rods and cones making up the picture elements of the eye's detection system. The inverted images collected at the retina are transferred to our brains via optical nerves where various abstractions are used in order to store the image. The exact colours, shapes and textures within the image are replaced with more familiar concepts such as 'dog' and 'light brown', greatly reducing the quantity of data required to recall a specific scene. Incidentally, the conceptual manner in which the brain stores images can be 'fooled' by simple tricks such as optical illusions. Today's computers do not have the intelligence to perform conceptual extraction and therefore basic methods are used for data storage.

There are two quite distinct methods of storing images within computer files: object (or vector) based and bitmap (or raster) methods. The object-based method describes images in terms of primitive shapes, i.e. a line in an image could be described by a position vector, its orientation, thickness and colour. By contrast, bitmap-based files store the image as a sequence of pixel intensities. In digital imaging, the bitmap method of image storage is the obvious choice because images from CCD cameras and other such devices usually capture the image in bitmap form. The object storage method is usually associated with graphic design packages where the image is constructed from sets of simple objects, such as squares and circles, each of which are described geometrically. Incidentally, an object-based image must be converted to a bitmap at some point if it is to be displayed on a computer screen or sent to a printer.

The structure of computer data files is complex and varies between operating systems. Files include not only the data but also information that is important to the computer operating system (OS). For example, this information includes the nature of the data (e.g. image, text, or some other document), the software needed to open the file or author information. However, these components of the

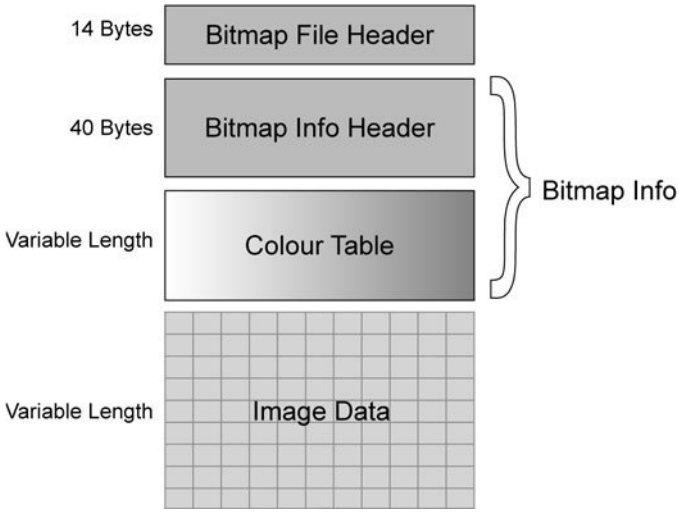


Figure 2.29 The structure of a Microsoft Windows™ bitmap image data file, where the first 14 bytes constitute a bitmap file header and the next 40 bytes contain the bitmap information header details. Finally, colour table information is provided before the pixel intensity data.

file are handled by the operating system so are normally of little concern to the user.

Ignoring the OS specific data, the structure of an image data file is split into two sections, the header and the image data. To illustrate the relationship between these two structures, a monochrome Microsoft Windows™ bitmap will be used as an example. A schematic of the Windows bitmap structure is shown in Fig. 2.29. The Bitmap Info structure contains two sections, a header section and a colour table. The Bitmap Info Header contains the data shown in Table 2.4.

The Bitmap Info Header is optionally followed by a variable length colour look-up table (LUT) which is used to specify the colours used in a 256 colour or less bitmap. In the case of 'true colour' images, the LUT is absent and the RGB values of each pixel are stored. The Bitmap Info Header is followed by the raw image data.

The Windows bitmap is a relatively simple image format and, for this reason, its header is concise. However, if an application were presented with the raw image data without the header information it would be unable to determine its dimensions and colour depth effectively rendering the image completely unusable. Any computer application which uses an image of a specific format must be able to read and interpret the information held within the header in order to be able to make sense of the raw image data. The structure of the header varies considerably between image types; however each has their own strict format and will usually contain the more common elements found in the Windows bitmap format illustrated above.

Table 2.4 Bitmap Info Header of a Microsoft Windows™ bitmap

Variable	Description
biSize:	The size of the Bitmap Info Header structure
biWidth, biHeight	The dimensions of the image in pixels
biPlanes	The number of planes of the target device, this must be set to '1'
biBitCount	The number of bits per pixel, ranging from 1 for a monochrome image to 32 giving a maximum of 2^{32} colours
biCompression	The type of compression (if any) used to store the image
biSizeImage	The size of the image in bytes
biXPelsPerMeter, biYPelsPerMeter	The dimensions of the image in pixels per meter on the target device
biClrUsed	The number of colours used within the image
biClrImportant	The number of colours required to display the bitmap

2.7.2 The choice of image format

Images are captured digitally for two broad purposes:

1. For presentation to a human user, where the exact quantitative details of the image are not important and only the main features of the image are required.
2. Analysis: to make measurements or to obtain some specific information held within the image. (This could range from simple quantitative measurements, such as the determination of the length of certain features, through to complex image processing algorithms, which involve many stages.)

In the former case the image must store enough information in order to recreate faithfully the image for the human eye. The detail required might be reduced when one considers the way in which the brain interprets images as described earlier. In the latter case the requirements of the image vary considerably, depending on the analysis which is to be performed. For example, many banks exchange and store cheques as digital images.[13] The most important parts of these images are the printed and handwritten text, which is bold and usually black. Of far less importance are the background patterns, which are printed on the cheque and which are usually more detailed and lighter. These patterns hold much less information about the nature of the cheque. Therefore these characteristic features can be taken into account when a suitable format is chosen for the storage of the images.

The data format of an image includes a number of elements such as colour depth and dimensions, which have both been discussed before. These will have an effect on the size (in bytes) of the bitmap image and hence the amount of memory or disk space required. For example, if a typical image is 1024×1024

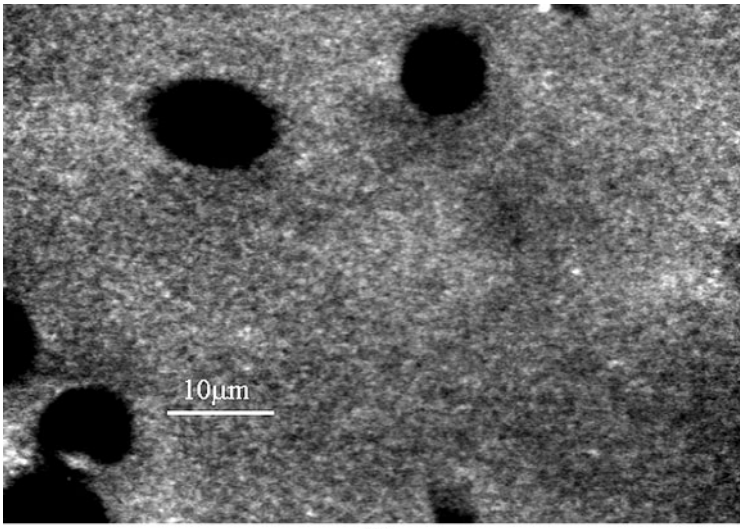
pixels in size and has a colour depth of 24 bits per pixel, the bitmap required for the storage of this image would occupy 3 MB of storage space. The fixed storage device found on most computers – the hard disk drive – typically holds 20 GB of data, enabling the storage of approximately 6000 of these images. If images are being captured on a regular basis, this figure could prove quite restrictive. Furthermore, when images are transferred between computers using portable storage devices or network connections, the restrictions of storage and bandwidth are even greater. In order to avoid these problems images are frequently compressed to reduce the strain on storage space and bandwidth.

There are a large number of compression techniques which can be applied to reduce the size of an image; for an extensive review of the techniques currently available readers are referred to Murray and van Ryper.[14] Whenever a compressed image is retrieved for displaying on the screen, printing or further processing, it must first be decompressed and stored in its bitmap format once more. At this point, any loss in image quality or artefacts introduced by the compression technique will become apparent. Compression techniques can be classed as lossy or lossless depending on the effect that they have on the image data.

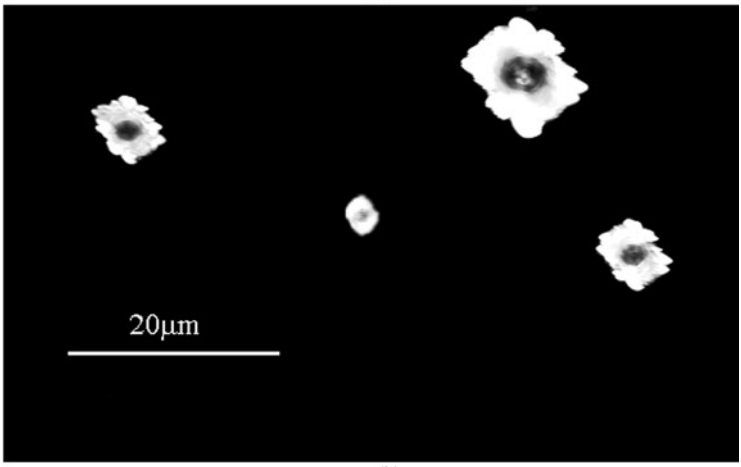
2.7.3 Lossless compression

The raw data that describes a bitmap image often contains long sequences of bytes that are the same. For example, an 8-bit greyscale image might have a sequential block of 23 pixels all of which have the colour '186'. Rather than storing the 23 repetitions, it is much more compact to store the intensity value '186' together with its number of repetitions. This process can be repeated throughout the entire image with the raw image data translated into a sequence of intensity values followed by their repetitions. For images containing a number of regions of uniform intensity or colour this description of the image data should be more compact than its respective raw format. In order to decode the compressed image, a program simply expands each pixel's intensity by the number of specified repetitions. This means of compression is a common one, and it is termed run-length encoding (RLE).

The performance of a compression algorithm can be evaluated by calculating the compression ratio, which is the ratio of the size of the original data compared to that of the compressed data. The performance of the RLE compression algorithm is highly dependent on the nature of the image. Those with large blocks of colour compress very well, whereas ones that have a lot of pixel to pixel variation such as noisy images are poorly compressed (see Fig. 2.30). In such cases more advanced compression algorithms such as substitutional compressors (which record repetitions of variable sized blocks of data as opposed to single bytes) may provide an improved compression ratio.



(a)



(b)

Figure 2.30 The performance of image compression routines is often dependent on the nature of the image. In this example, two images have been compressed by run-length encoding. (a) A noisy image can be compressed only by 34%. (b) An image, with large regions of similar colours, may be compressed by 91%.

2.7.4 Lossy compression

It is not always necessary to reproduce exactly the information found within the image data; it may be sufficient to replace patterns within the image with a more generalised version for which there is a more concise description. However, when the image is decompressed the resultant image will not be the same as the original image, hence this technique is known as lossy compression.

One of the most widely used lossy compression techniques, due to its widespread use on the Internet, is the Joint Photographic Experts Groups (JPEG) standard.[15] The JPEG compression algorithm is a two-stage process. Firstly, the image is split into colour and brightness information. The spatial resolution of the colour information is reduced by a factor of two, because the human eye is more sensitive to changes in brightness than colour. The second stage inspects 8×8 blocks of pixels and looks at the spatial frequency within this region. The human eye is less sensitive to higher frequencies and therefore these are removed from the image. The amount of frequency information removed from the image determines by how much it is compressed, giving a convenient tool for varying the compression ratio. The combined reduction of colour information and higher frequencies results in a much more compact description of the image.

The reasoning behind the JPEG compression methodology has evolved because of our understanding of how the human eye and brain perceive images. Therefore, the JPEG compression standard is most suited to applications where images are only intended for display or illustrative purposes. If the images compressed by the JPEG standard are used subsequently for image analysis, undesirable artefacts can be generated such as those demonstrated in Fig. 2.31. In this example, the edges of a filled circle are being detected using a 3×3 Laplace filter (described in a later section). After the application of the convolution filter to the original images, the edges are clearly visible. However, the application of the same filter to the JPEG compressed image shows a very poor edge profile.

2.8 Image enhancement

The objective of image enhancement depends on the intended application of the image (in much the same way as the choice of image storage format). If the image is intended for presentation to a human user, either on a monitor screen or in print, image enhancements may be made in order to ensure the features of interest are more clearly visible. Recently the introduction of digital cameras, scanners and inexpensive home computers has made the enhancement of digital photographs commonplace. Techniques such as contrast enhancement, red-eye removal and photograph restoration are frequently used. If the image is intended for further processing, for example segmentation and feature extraction, image enhancement can improve the success of subsequent processing steps. All the operations have in common that they start with an image and return another (transformed) image.

The image enhancement techniques that follow are roughly divided into four sections. Firstly, corrective processing routines are described (and these routines are certainly of great importance, if the result is to be unbiased). Secondly, simple methods for histogram manipulation will be considered. Thirdly, a

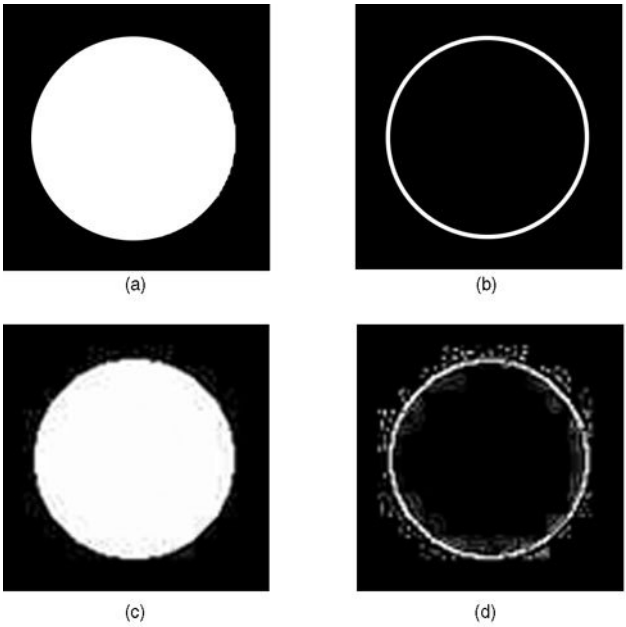


Figure 2.31 The lossy effects of JPEG compression are acceptable when the image is used for visualisation purposes only. However, in this example, where the image is subsequently processed, the artefacts of JPEG compression are enhanced to a point where they are unacceptable. (a) The original image, and (b) the result of applying an edge detection algorithm. (c) The JPEG compressed image, and (d) the result of applying an edge detection algorithm on the compressed image.

number of spatial filters will be described and various applications highlighted. Finally, a brief overview of frequency-domain filters will be described.

2.8.1 Corrective processing

Ideally the image of an object should be spatially invariant, i.e. its intensity, shape and size should remain constant as the feature is moved within the field of view. The optics of a modern microscope is corrected in order that this may be achieved. However, incorrect usage of the microscope can degrade the image quality, for example, uneven sample illumination can result in an imaging system where objects change in appearance as they are moved across the field of view. A microscope operator, equipped with a highly advanced image processing system (the brain), would probably realise that the change in the object's appearance is due to the non-uniform illumination. However, a computer-based image processing algorithm may not realise this fact. Hence, characterisations of the same object will change, depending on its position in the image. This is clearly undesirable and illustrates the important point that correct

usage of the microscope is of utmost importance when the images are being processed by a computer.

The path an image takes from the light reflected by the sample to its final destination as a digital image is long and involves storage/transfer in a number of different formats. It is inevitable that the final image will not be a perfect, spatially invariant image. Fortunately, there are many image processing tools which can be applied to the image in order to correct some of these defects in the final image.

The first technique is frame averaging. Images (or any other data) acquired from modern sensing equipment are susceptible to noise. The observed image is composed of two components: the true signal, and an added amount of random noise superimposed on the signal. The term noise refers to random variations in signal intensity as opposed to distortions such as shading or lack of focus. The image is transformed a number of times during its path from the microscope to its eventual digital format inside the computer. At each stage, noise may be introduced simply by the digitisation process itself (see Section 2.5.2). While modern technology has made it possible to reduce the noise levels, it is unreasonable to assume that they will be eliminated altogether.

Since this noise component is statistically random, its effect on the image can be reduced by frame averaging. Accumulating N images will increase the intensity of the noise component by the square root of N. However, at the same time, the intensity of the true signal will increase by N. Therefore, the signal-to-noise ratio in accumulated images will increase by the following factor.

$$S_N = \frac{N}{\sqrt{N}} \qquad [2.18]$$

The reduction in noise achievable by accumulating image frames is shown in Fig. 2.32.

The second technique for improving the raw image is background correction. One of the most common forms of defects that creep into the image during digital capture is shading. This frequently takes the form where there is an intensity gradient across the image caused by uneven sample illumination. However more complex non-uniformities in illumination can occur as illustrated in Fig. 2.33, where the non-linear response of an acousto-optical deflector, used in a confocal laser scanning microscope, has resulted in a highly uneven background intensity.

The easiest approach to dealing with this problem is to calibrate the imaging system by collecting a reference background image. The method used to acquire this image depends on the imaging mode being employed. For transmission microscopy a background image can be acquired when the field of view is empty which effectively images the light source on the focal plane. In reflected light microscopy a suitable test specimen for collection of a background image must reflect the light illuminating evenly throughout the field of view. A mirrored surface is suitable for this purpose, however any scratches or defects in

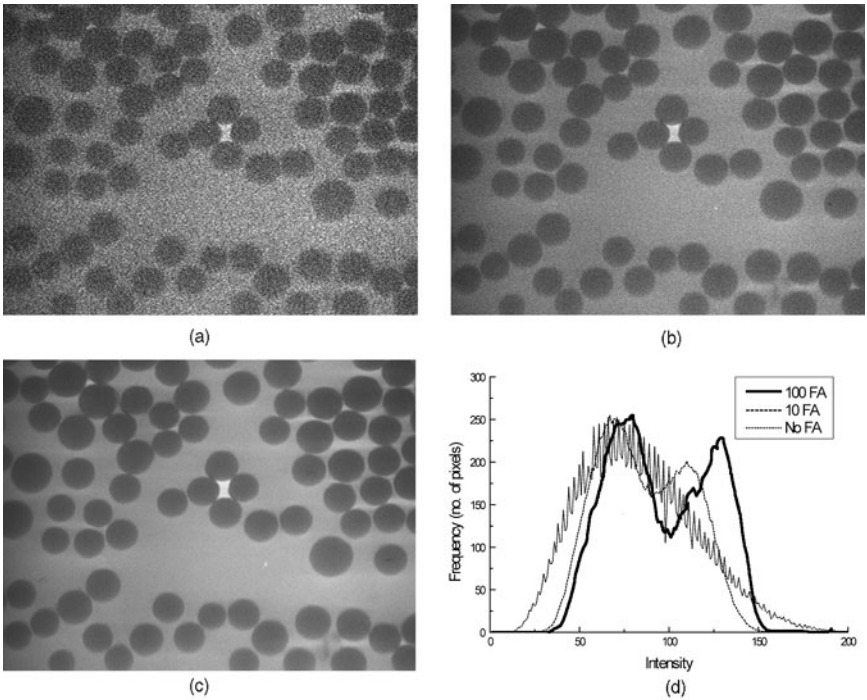


Figure 2.32 A number of images of the same region demonstrating how frame averaging improves the quality of the image. (a) A single frame, (b) 10 frame averages, (c) 100 frame averages, (d) pixel intensity histograms are overlaid for the three cases. The images shown have an area 100 μm$\times$75 μm.

the mirror will also be stored in the reference image. This can be minimised if a number of images are collected, with the slide moved slightly between each capture. Averaging all of these images can form a reference image.

The suitable method for correction of the background intensity variation depends on the source of the error. If the source of the shading is due to a permanent offset in image intensity then subtraction is the best method of correction. If the shading is caused by attenuation of the light illuminating or emanating from the sample then division is the best method of correction. Frequently, a combination of both is observed.

Equation [2.19] shows the correction calculation, which is evaluated for each pixel of the image. $I_o(x,y)$ is the original image intensity, $I_c(x,y)$ is the corrected intensity, $B_r(x,y)$ is the reference background image, $B_e(x,y)$ is an image acquired of the field of view when the illumination intensity is zero and $< I_o >$ is the mean intensity of the original image.

$$I_c(x,y) = \frac{I_o(x,y) - B_e(x,y)}{B_r(x,y) - B_e(x,y)} < I_o > \qquad [2.19]$$

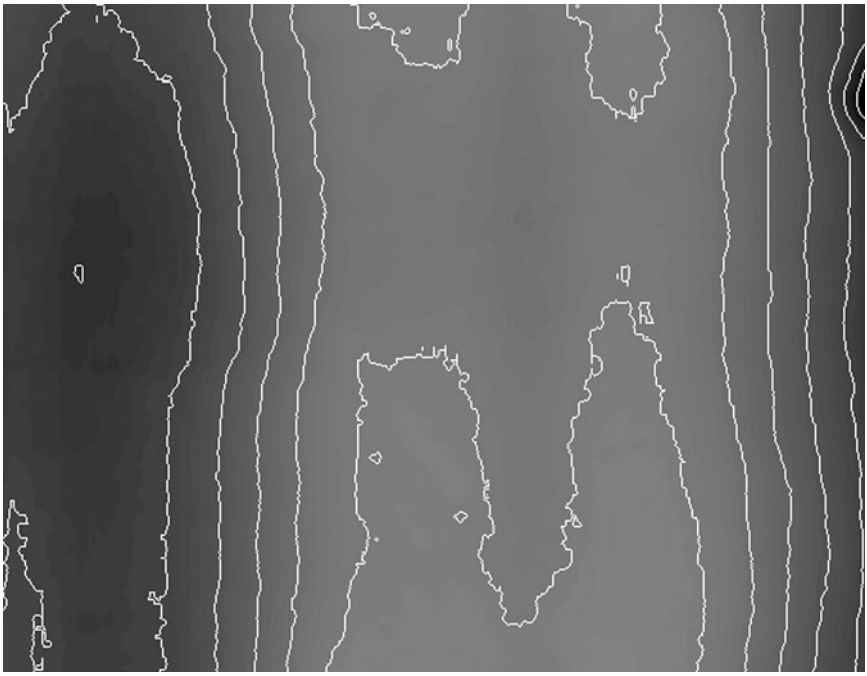


Figure 2.33 A reflection mode image of a mirror clearly demonstrates the non-uniform image gain over the field of view. Intensity contours are plotted for pixel intensity values separated in steps of 10 units.

The third technique for improving the raw image is geometric correction. One of the first steps that needs to be taken before making quantitative measurements is the calibration of the field of view. This usually involves the capture of a few digital images of a suitable calibration slide, from which the scale of the individual pixels can be measured. However, if the image has been distorted at all during its capture, it will be quite apparent because calibration measurements taken from different parts of the image will give different results. The presence of a geometric distortion can result in serious problems when the images are processed further, for example it will not be possible to find a good match between neighbouring images.

Geometric distortions, which include stretching and skewing are usually smooth. A smooth transformation of an image can be described by an equation which relates the location of a pixel in the target image $t(x,y)$, to a location within the source image $s(x,y)$ via a pair of functions $i(x,y)$ and $j(x,y)$.

$$t(x,y) = s(i(x,y),j(x,y)) \qquad [2.20]$$

If the two functions $i(x, y)$ and $j(x, y)$ are known, or determined, it is possible to reverse the transformation in order to recover the non-distorted image. In

practice it is unlikely that the characteristics of the geometric distortion will already be known, therefore these must be determined experimentally.

2.8.2 Histogram manipulation

The histogram of an image is a frequency plot, which shows the number of pixels at each intensity level and provides a global description of the appearance of the image. The technique of histogram manipulation uses a monotonic transformation function to map the intensity of each pixel (within the original image) into a new pixel intensity. A modified intensity histogram therefore results. This is a point transformation, in which the resulting pixel intensity depends only on the original pixel intensity and the transformation function. By contrast, the spatial and frequency filters described later affect each pixel intensity depending on the values of neighbouring pixels.

The transformation of an image histogram is illustrated in Fig. 2.34. The transformation function is continuous, but the histogram of a digital image is a discrete function because pixel intensities are most often stored as integer values. As each intensity is transformed, it must be rounded to the closest acceptable discrete intensity, which may result in a fragmented intensity histogram (see Fig. 2.34).

Almost any transformation function can be used, although there are some which are more commonly applied than others. Figure 2.35 shows three examples alongside the original image: logarithmic, inversion and histogram equalisation. The latter transformation is of particular interest because the image is transformed to give an equal frequency of each discrete intensity level (or as

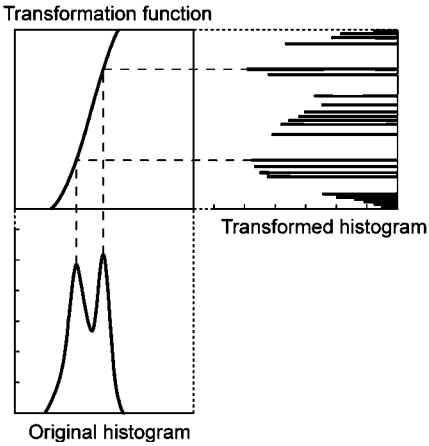


Figure 2.34 The histogram of an image before and after the application of the histogram equalisation transform function. Note that the transformed histogram is fragmented.

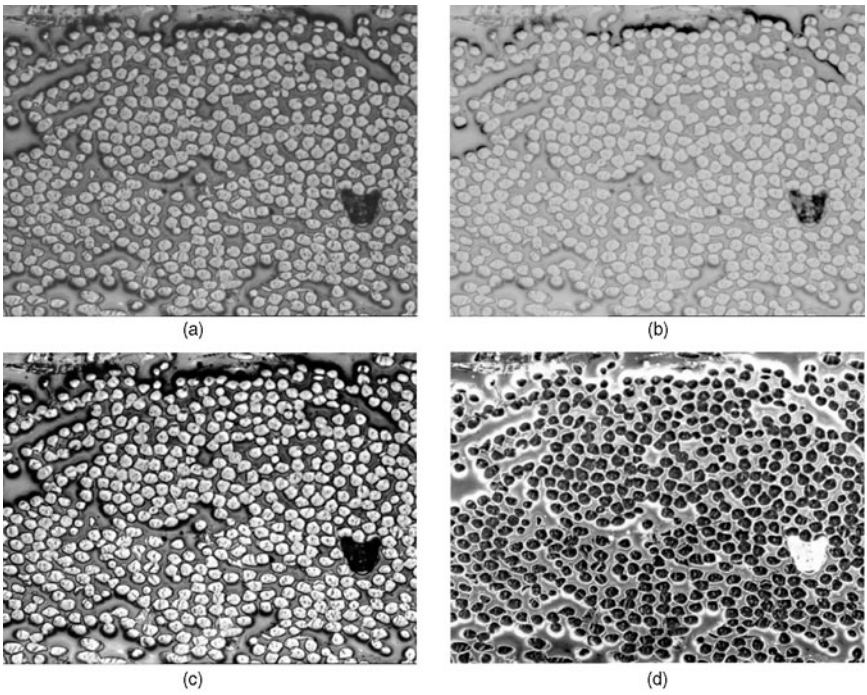


Figure 2.35 A number of common histogram manipulation routines applied to the same fibre-reinforced composite image. (a) The original image, (b) after logarithmic histogram stretch, (c) after inversion and (d) after histogram equalisation. Each image represents an area of approximately 150 μm$\times$100 μm.

close as possible, due to fragmentation). The histogram equalisation transformation function is derived from the original image and the details of this procedure can be found elsewhere (e.g. Gonzalez and Wintz[16]).

A greyscale image may be false-coloured by a transformation function with separate red, green and blue components. This technique is especially useful for accentuating details in an image which is being displayed. The human eye can only distinguish approximately 30 grey levels, which is far fewer grey levels than the majority of modern computers are capable of displaying. However, the human eye should be able to distinguish 30^3 different colours and therefore a false-coloured image provides more information to the human viewer.

The manipulation of an image's pixel intensity histogram does not really change subsequent image measurements. Furthermore, the fragmentation of the transformed image histogram results in a non-recoverable loss of data. Therefore, if the image is not intended for display purposes, it is much more desirable to modify the subsequent image processing routines so that they function properly on the non-modified image.

2.8.3 Spatial domain filters

Spatial, or convolution, filters are applied to an image by considering each individual pixel in turn. However, they differ from the point transformations described earlier in that the intensity values of the local neighbourhood affect the transformed intensity of an individual pixel. The convolution of a single point (x,y) within the image $f(x,y)$, with the $s \times s$ kernel, $h(x,y)$, is stored in the new image $g(x,y)$ as follows:

$$g(x,y) = c_h \sum_{i=x-s}^{i=x+s} \sum_{j=y-s}^{j=y+s} h(x-i, y-j) f(x,y) \qquad [2.21]$$

where c_h is a constant. This process is repeated for all the pixels (x,y) within the original image until the entire image array $g(x,y)$ is filled.

Each element of the kernel is a weighting factor and these weights determine the effect that the particular convolution will have on the resulting image. For example, consider the neighbourhood average type filter,

$$h(x,y) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad [2.22]$$

This filter will replace the value of the central pixel with the mean value of all of the neighbourhood pixel intensities.

Some of the more common 3×3 convolution kernels will be mentioned together with their effects on a test image. One typical 3×3 convolution filter is a Laplace transform, which is frequently used to locate the edges of objects, as illustrated in Fig. 2.36. When these convolution filters are applied in practice, two things must be considered.

Firstly, pixels at the edge of the image do not have the full complement of neighbouring pixels (making it impossible to apply equation [2.21] as defined above). There are a few possible solutions to this problem, for example, the edge pixels could be ignored, making the transformed image slightly smaller than the original one. Alternatively, extra pixels could be fabricated and assigned a suitable intensity. Another possibility would be to crop the kernel so that it always fits within the physical dimensions of the image.

Secondly, the pixel intensity, determined by equation [2.21], may not always be within the dynamic range of the image type being used. The constant c_h is intended to 'push' the resulting image into the correct intensity range. For example, the 3×3 kernel, equation [2.22] where each of the weights is '1', will have the effect of increasing the overall intensity of the image. (This is because each new pixel intensity will be equal to the sum of its nine neighbours.) However, it is easily shown that the mean intensity can be maintained if the result of each summation is multiplied by the constant, $c_h = 1/9$.

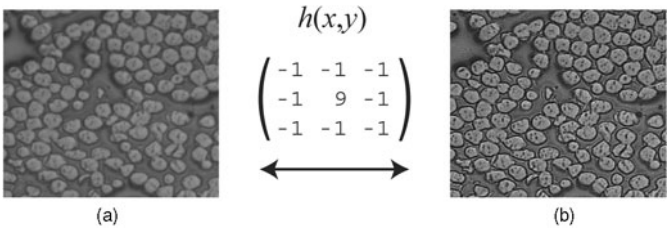


Figure 2.36 (a) The original image of a fibre-reinforced composite is shown. (b) The resultant image after each pixel in the original image has been processed by the 3×3 convolution filter, which enhances the edges within the image.

2.8.4 Frequency domain filters

The spatial filters described earlier are so called because they work within the spatial domain, i.e. the transform acts directly upon the intensity values within the image. Another important class of filters is image filters, which work in the frequency domain. The filters described here all utilise the well-known Fourier transform, although this is by no means the only method available. Similar filters may be applied to the discrete cosine spectrum, for example.

The continuous form of the Fourier transform $F(u,\nu)$ of an image $f(x,y)$ is as follows:

$$F(u,\nu) = \int^{-\infty} \int_{-\infty} f(x,y) e^{-j2\pi(ux+\nu y)} dxdy \qquad [2.23]$$

and the discrete form for a square image of size N pixels is as follows:

$$F(u,\nu) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi[(ux+\nu y)/N]} \qquad [2.24]$$

where $u,\nu = 0,1,2,\ldots N-1$. The transformed image F contains all the information of the original image, f, which can be recovered by application of the inverse Fourier transform. The frequency space F is complex, containing both real and imaginary parts. For purposes of illustration, the magnitude of F is usually plotted with $F(0, 0)$ at the centre of the image. The presence of low frequencies within the image result in peaks near the origin, $F(0, 0)$ and higher frequencies result in peaks progressively further away from this point.

The calculation of the discrete Fourier transform involves N^4 complex calculations and, therefore, computation time increases considerably as the image size is increased. Most image processing software employs an optimised Fast Fourier Transform (FFT) which dramatically reduces computation time. There are many texts which cover the design and optimisation of the FFT algorithm and a good starting point is the 'Numerical Recipes' series published by Cambridge University Press (for example Press *et al.*[17]).

The equation for the convolution of an image by a kernel, $h(x,y)$, in the spatial domain was given in the previous section. Exactly the same convolution can be applied in frequency space as follows:

$$G(u,\nu) = H(u,\nu) \cdot F(u,\nu) \qquad [2.25]$$

where F, G and H are the Fourier transforms of the original image, resultant image and the convolution kernel respectively. Ignoring the time taken to compute the FFT, this is a much simpler operation than its spatial domain counterpart. Image processing routes frequently involve the application of a number of convolution filters. If many images are being analysed, or real time processing is required, processing in the frequency domain would certainly yield superior results.

Figure 2.37(a) shows an image of a mandrill baboon face and Fig. 2.37(b) the Fourier spectrum associated with the image. This test image contains regions of both high frequency information (the long hairs about the face) and low frequency information (the region about the nose and eyes). Regions of the Fourier spectrum can be removed or diminished in order to suppress features within the original image. Figure 2.37(c) is an illustration of high-pass filtering of the same image, where the circular region within the Fourier spectrum was removed. Although the fine details within the original image are still present, the lost lower frequency information, which provides the background shading, produces a poor quality image. Figure 2.37(d) shows an example of low-pass filtering, where the information outside the marked circle was removed. In the resultant image, the regions within the original image, which were mostly composed of low frequencies (the nose and eyes), are quite well preserved. However, the fine details in the image are completely lost. Returning once more to the subject of compression, this example shows how the removal of higher frequencies in a region dominated by low frequency can yield an image which is still acceptable to the human viewer.

Fourier spectrum methods are especially useful for the analysis of images that contain periodic structures. Figure 2.38(a) shows a regular array of $5\,\mu m$ latex spheres. The sample contains a small number of defects, which lead the eye away from the underlying regular structure. Inspection of this image's Fourier spectrum (Fig. 2.38(b)) shows a radial band of peaks, which correspond to these regular structures. The spectral components outside this band were removed and an inverse Fourier transform applied, resulting in Fig. 2.38(c). Clearly, the latex spheres are enhanced considerably in the resulting image. Furthermore, the spacing and orientation of the periodic structure can be measured directly from the Fourier spectrum. Conversely, Fig. 2.39 shows an image which has been affected by periodic interference. Elimination of the corresponding peaks in the Fourier spectrum provides a very effective method for the removal of this type of interference.

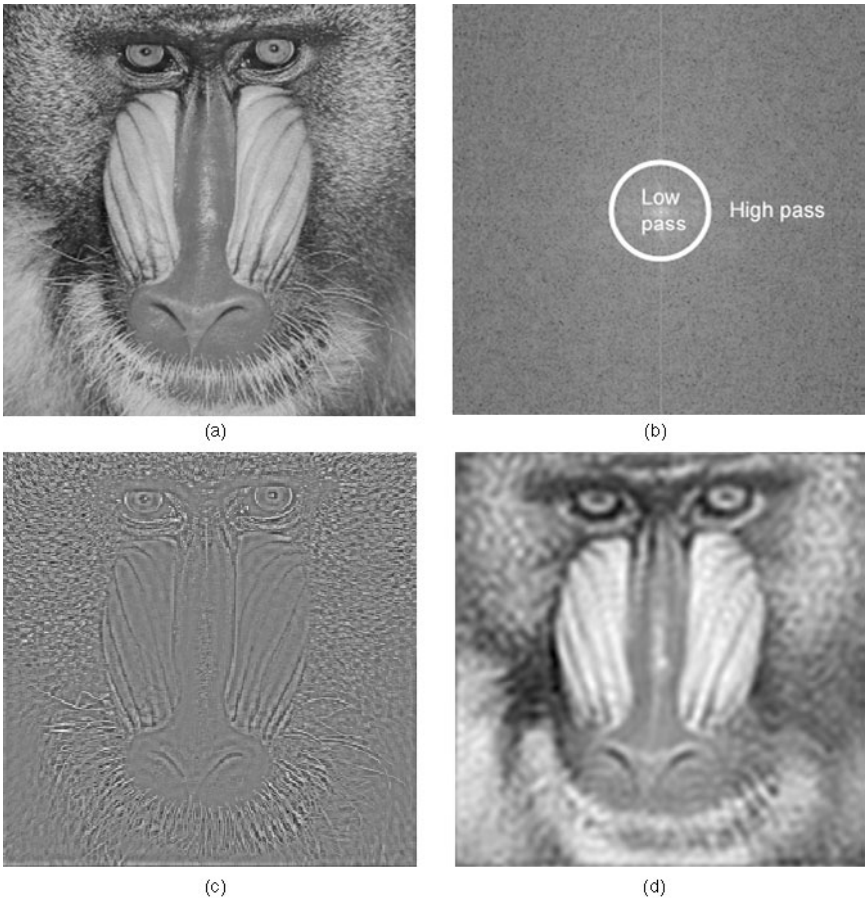


Figure 2.37 (a) A digital image of a baboon and, (b) its Fourier transform. In (c) and (d) the image's Fourier spectrum has been high-pass and low-pass filtered, respectively.

2.8.5 The Hough transform

In 1962 Paul Hough patented a transform for 'machine recognition of complex lines in photographs and other pictorial representations'.[18] The Hough transform is now widely applied to problems in image processing and machine vision for the detection of straight lines, circles and other curves. This transform is particularly robust to missing and noisy data, which has contributed to its rise in popularity.

As an example, consider an image containing n pixels that have been identified as object boundaries. If we wish to determine subsets of these pixels, which lie on straight lines, one possible approach would be to test all the possible straight-line connections and determine how many pixels lie along each. However, for all but the most trivial of cases, there are a huge number

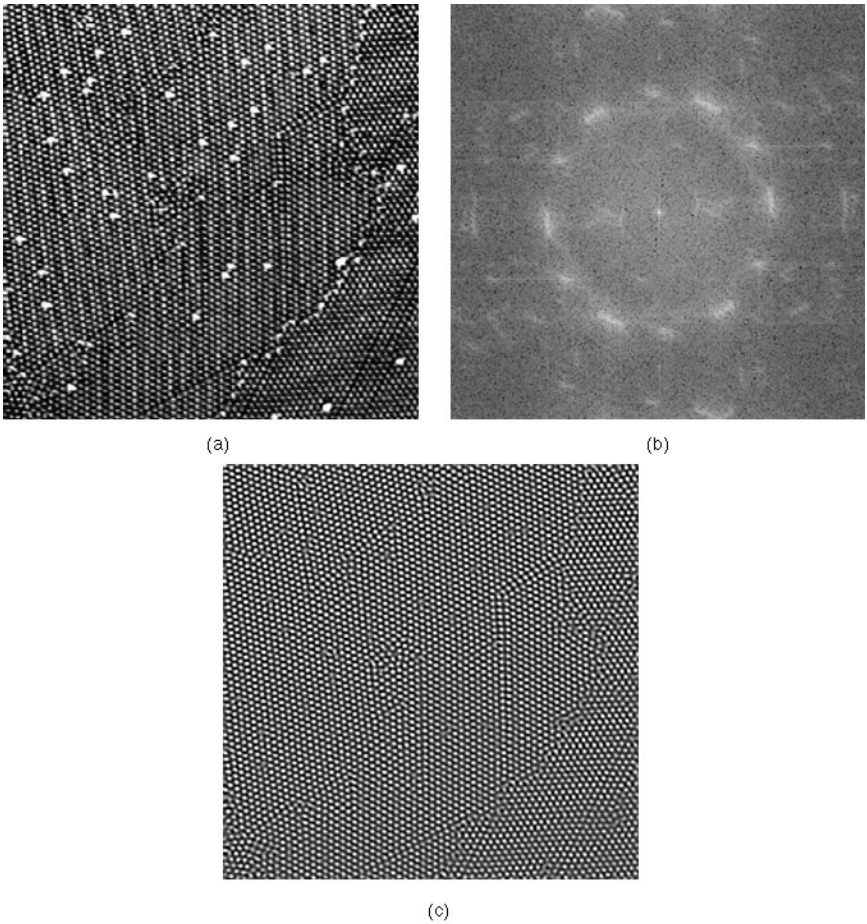


Figure 2.38 The Fourier spectrum of an image can be modified to enhance periodic features. The sample illustrated in (a) is a matrix of 5 μm latex spheres dispersed using a solvent, which has evaporated. (b) The image's Fourier spectrum is shown where the periodic nature of the image is seen as a circle of peaks about the centre. (c) These peaks have been enhanced and a cleaned-up version of the original image has been created.

(approximately n^2) of possible lines to check. The Hough transform approaches this problem in a different way. Rather than considering lines connecting pairs of pixels, each individual pixel is considered in turn and *all* the possible lines passing through this point are determined.

A straight line is described in parametric form by the following equation:

$$x \cos \theta + y \sin \theta = \rho \qquad [2.26]$$

where ρ is the length of a normal from the origin to this line and the angle, θ, is the orientation of the line. In the above problem, the edge pixel coordinates

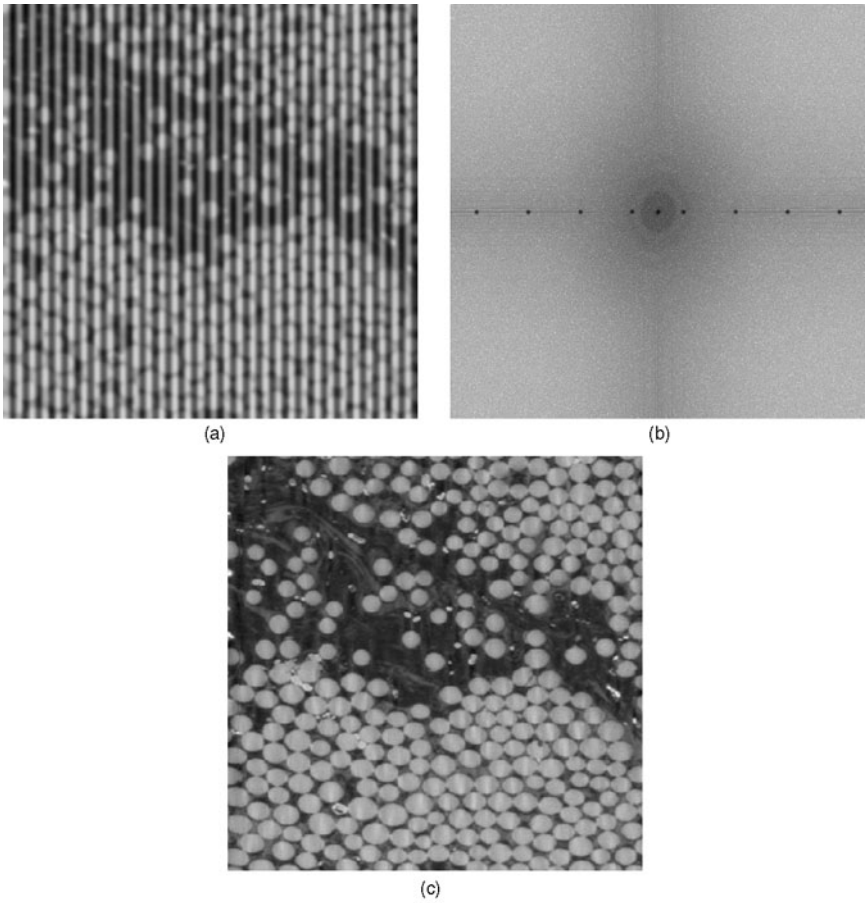


Figure 2.39 (a) An image which has been subjected to regular spatial noise. (b) The two peaks in the Fourier spectrum may be suppressed. (c) By suppressing the peaks and recreating the image with an inverse Fourier transform, the periodic noise is removed.

(x_i, y_i) are already known and serve as a constant in our equation, while θ and ρ are the values which we seek.

It is straightforward to determine the possible (θ, ρ) values for an individual edge pixel from equation [2.26] and these values describe every possible straight line which passes through the point (x_i, y_i). A plot of these values in the Hough parameter space forms a sinusoid. Figure 2.40(a) shows three pixels within an image and Fig. 2.40(b) shows their respective sinusoids in Hough parameter space. The three pixels in this example are collinear, therefore there is a unique line, (θ_i, ρ_i), which passes through all three. Therefore, the Hough space sinusoids for each pixel will all pass through the point (θ_i, ρ_i) and hence they

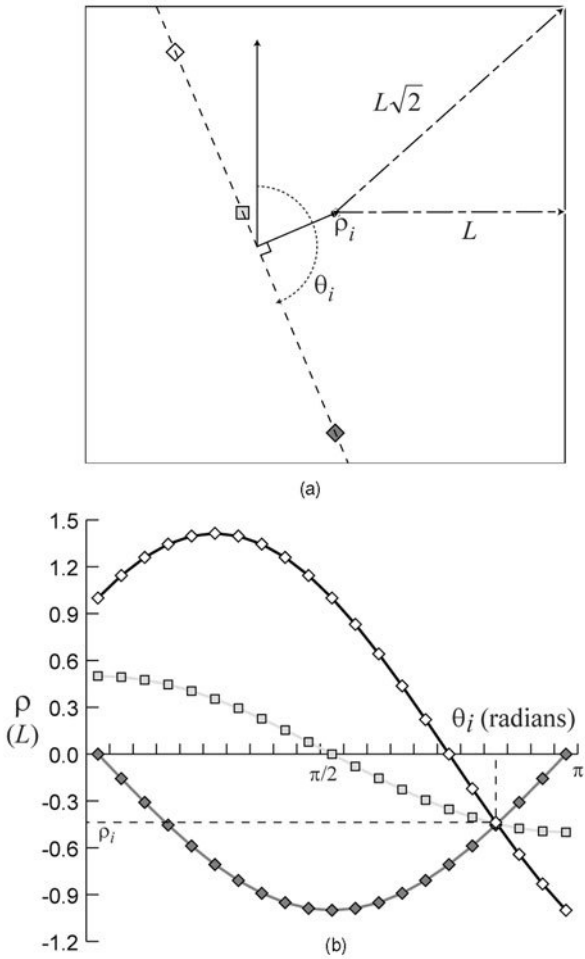


Figure 2.40 An illustration of the linear Hough transform. (a) Three collinear points have been identified in the image plane. (b) When the Hough transform is applied to all of the lines *which could pass* through each of these three points, three curves are generated in the parameter space (ρ, θ). Their point of intersection in the parameter space defines the ρ_i, θ_i of the line that joins all three collinear points.

will coincide. In an image containing a large number, n, of edge pixels, collinear pixels can be located by searching for intersections in Hough space.

Equation [2.26] is a continuous function giving an infinite number of (θ, ρ) pairs for a single image pixel. In practice, the parameter space is quantised into finite intervals and, for a linear Hough transform, it is a two-dimensional array. This accumulator array is initially empty. As the algorithm runs, each image pixel is transformed into a discretised (θ, ρ) curve and the accumulator cells in

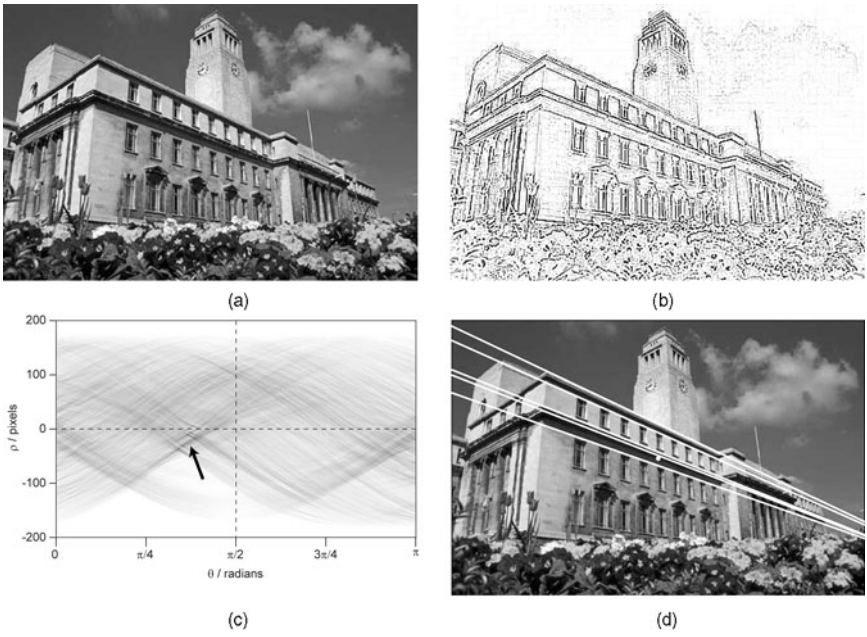


Figure 2.41 An illustration of line detection using the Hough transform. The edges of features within the original image in (a) are detected using a 3×3 Laplace transform with the results shown in (b). The Hough transform of this image is shown in (c) and the 'peaks' or dense regions located by the arrow, within the parameter space, reveal the strongest lines within the original image, as illustrated in (d).

the parameter space which lie along this curve are incremented. Sinusoids which coincide, as illustrated in Fig. 2.40(b), result in peaks in the accumulator array and they represent strong evidence that a corresponding straight line exists in the image.

Figure 2.41(a) shows a digital image of a building (the Parkinson Building, University of Leeds) and the edges have been detected using a 3×3 Laplace convolution filter, as illustrated in Fig. 2.41(b). Each of the pixels within the edge-detected image, which are above a prescribed threshold intensity, have been transformed into curves in parameter space. The value of each cell in the accumulator array can be displayed as an intensity level, enabling the parameter space to be visualised, as shown in Fig. 2.41(c). A small cluster of peaks have been identified in the parameter space and their (θ, ρ) values recorded. Figure 2.41(d) shows the respective lines that these values represent in the original image. It can be seen that these peaks correspond to the longest lines present within the image. Smaller peaks can be identified within the parameter space in order to locate some of the shorter lines within this image.

The origin location in Equation [2.26] is arbitrary, as long it remains fixed throughout the entire transformation. One might intuitively place the origin in the top left corner of the image, in order to coincide with the origin of the digital image. Indeed, there are a number of image processing books which illustrate the Hough transform in this way. However, this places a restriction on the parameter space, with none of the sinusoids entering the top right quadrant resulting in a skewed plot. In the example illustrated in Fig. 2.41, an origin was chosen in the centre of the image giving an evenly distributed parameter space.

The Hough transform can also be used to obtain global image characteristics such as orientation distributions. For example, see the case study on X-ray micro-tomography in Chapter 5 and the textile work of Xu and Yu.[19] Although this section has focussed on the linear Hough transform, the same technique can be applied to curves. For example, a circle can be described by the following equation:

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \qquad [2.27]$$

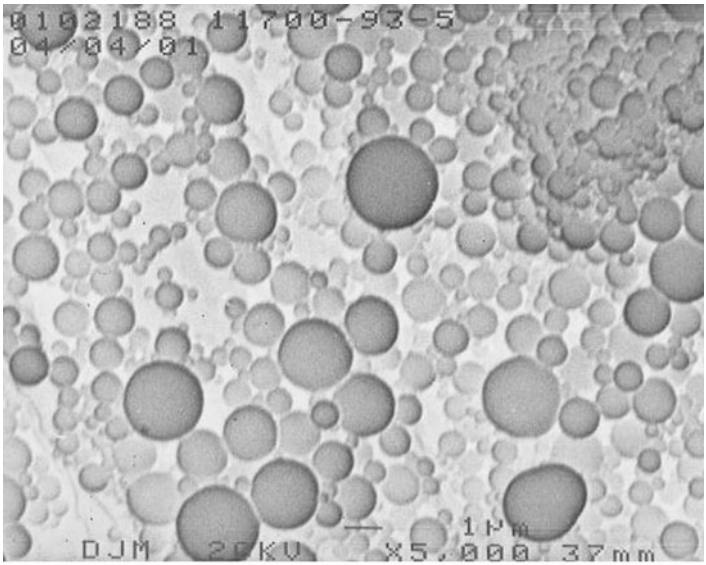
where r is the circle's radius and (x_0, y_0) are the centre coordinates. This equation has three parameters and, therefore, the parameter space is three-dimensional. Each point, (x_i, y_i), in the image is transformed into a cone in parameter space. The application of a circular Hough transform to an image of soap bubbles is shown in Fig. 2.42(a). A cross-section of the 3D parameter space, for a constant radius, is shown in Fig. 2.42(b).

From the above examples of linear and circular Hough transforms, the reader may be led to believe that the Hough transform can only be applied in situations where a simple analytical description of the feature can be found. However, in those cases where this cannot be achieved, the generalised Hough transform can be applied, in which the feature is described by a look-up table of distance and direction pairs.[20]

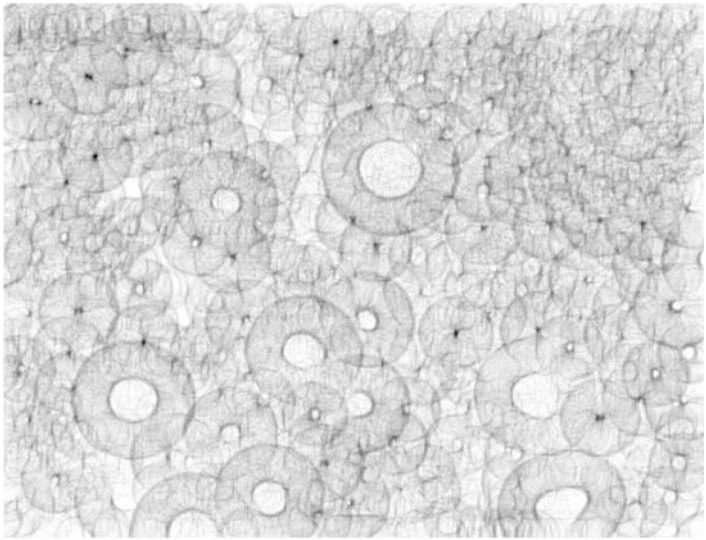
2.8.6 Segmentation

The image processing routines described so far have been used to enhance the image, or features present within the image. In all cases the routine results in another image, i.e. it tells us nothing about the nature of the image and does not give information about features within the image. However, appropriate image enhancement is always chosen with *a priori* knowledge of the nature of the images. For example, the Fourier spectrum modification applied to the image in Fig. 2.38 was chosen specifically to enhance the periodic features, which have already been identified within the image. The next logical step is the identification of the objects within the image, separating them from the background. This process of segmentation is one of the most important steps in any image processing routine.

Many image properties can be used to segment an image. The easiest, and possibly the most widely used, method is to segment regions by their relative



(a)



(b)

Figure 2.42 Circular features within images can be identified using a circular Hough transform. The bubbles in image (a) are detected by locating their peaks (the dark spots) in the parameter space, which are illustrated in (b).

intensity. The careful preparation of a sample, which usually involves cutting and polishing, is essential in order to accentuate these intensity differences. Microscopists working in materials science are perhaps more fortunate than their counterparts working in biological sciences, who are often faced with samples

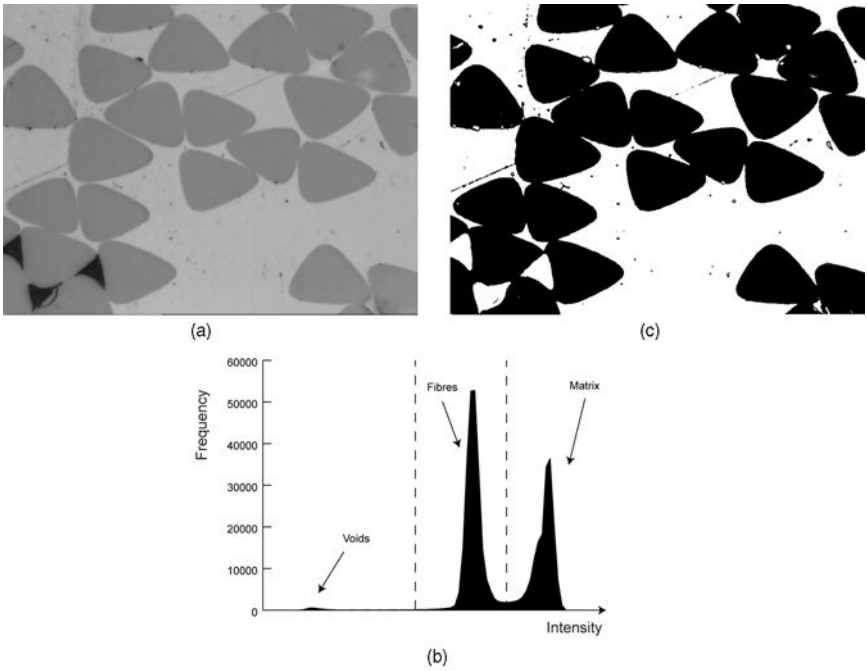


Figure 2.43 Segmentation involves splitting an image into different regions of interest. (a) The sample has been prepared carefully so that there is good contrast between the features. (b) When the frequency histogram of pixel intensities is inspected, the features are readily identified. (c) By defining an intensity threshold between the matrix pixels and fibre pixels, a binary image can be created to allow further image processing and hence characterisation of the fibres.

with poorly defined boundaries that even the human eye finds difficult to segment.

The segmentation of an image by its intensity requires the selection of a threshold (or if multiple regions are being identified, more than one threshold) which divides the image into the required regions. Hence, this operation is sometimes termed 'thresholding'. The selection of an appropriate pixel threshold intensity is often apparent on inspection of an image's intensity histogram. This is illustrated in Fig. 2.43, which shows an image of a fibre-reinforced composite and its accompanying histogram. The three regions visible in the image – voids, fibre and matrix – are clearly visible as separate peaks in the intensity histogram. If the fibres are of interest, the intensity range containing its respective peak is selected. All image pixels outside this intensity range are replaced with an intensity value of 0 (black) and all those inside this range are given an intensity value 1 (white). This procedure results in a binary image, which is segmented into regions of fibre and non-fibre.

With the image segmented into regions of fibre and non-fibre, it is possible to measure the relative density of each, however further processing is required in order to characterise the objects, fibres in this case, themselves. Objects can be defined loosely as regions of pixels which are connected. For instance, if two pixels are in the same object, a path can be found between them, and this remains within the object. This connectivity property can be used to identify objects by a process known as 'region growing'.

An iterative region-growing algorithm using a recursive function is described here. Firstly, a 'seed' point is selected anywhere within the chosen object. During each iteration of the algorithm, the neighbouring pixels of the 'seed' point are visited. These pixels are each compared against a 'membership criterion'. In this case, they must have an intensity of '1' to be members, although criteria that are more complex can be implemented. Any pixels passing these criteria also become 'seed' points and the entire process is repeated. By this technique, the region is 'flood filled' until all boundaries are reached. Figure 2.44 illustrates how a small region may be grouped by pixel connectivity. The first image shows the original seed pixel, the subsequent images show how the region grows with the black pixels indicating the growth fronts. When the boundary is reached in all directions, the region is complete.

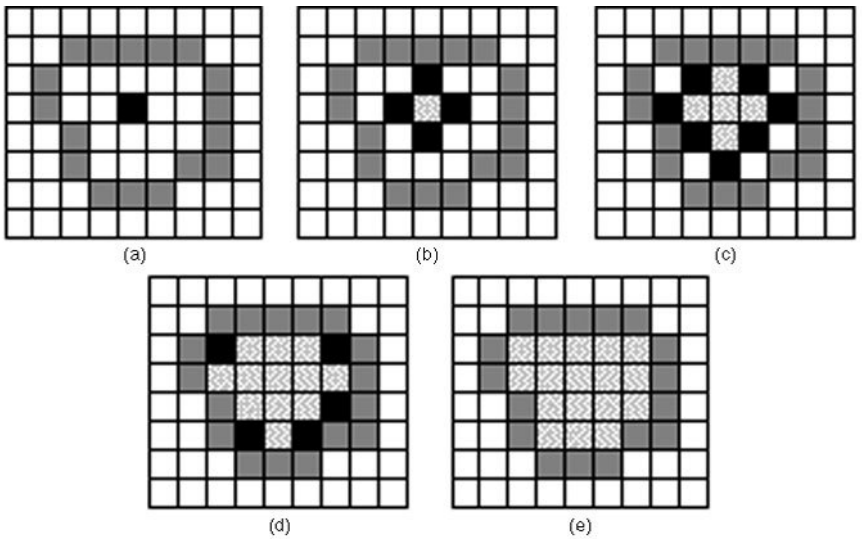


Figure 2.44 An illustration of the recursive pixel connectivity algorithm showing the iterative 'flood fill' of a segmented region. The images run sequentially from (a) through to (e), where the black pixels are 'active', i.e. they are the flow-front of the flood fill.

2.8.7 Characterisation and classification

After segmentation, the objects of interest within the image have been identified, however their nature has still to be clarified. Characterisation involves describing each object within the image by a small number of parameters and, hence, describes the overall image (and the sample) in a more concise manner. The choice of characterisation parameters is very much dependent on the nature of the sample and its particular application. For example, in the fibre length case study in Chapter 3, clearly it is desirable to characterise the features within the image by their length. However, in the fibre orientation case study in Chapter 3, the objects (fibre cross-sections) are characterised by their absolute positions and also their ellipticities (so that the fibre orientations may be calculated).

Finally, the decisions as to the nature of each object can be made by analysing trends in the various characterisation parameters. Again, using the fibre orientation case study as an example, the objects within the image are classified as either valid fibre cross-sections or 'other' objects (which are mostly fractured fibres due to the sectioning and polishing process).

2.9 Errors and stereology

> In his errors a man is true to type. Observe the errors and you will know the man, …
>
> Confucius, *Analects*, IV.7

The results of any scientific experiment, obtained using either a microscope or other laboratory apparatus, should always be accompanied with at least some reference to their accuracy. The need for these accompanying data can be illustrated in the following three simple sentences. It is impossible to make an exact measurement. Therefore, all experimental results are wrong. Just how wrong they are depends on the kinds of errors that were made in the experiment. This may sound like an unnecessarily pessimistic point of view, however, it is in line with the Heisenberg 'Uncertainty Principle' (one of the cornerstones of modern physics).[21]

The errors associated with any experimental measurement can be classified as either;

- *Random error* – these are chance variations in the measurements over which you as experimenter have little or no control. (Since the errors are equally likely to be high values as they are to be low values, averaging a sufficiently large number of samples will improve, in principle, the estimate of the 'true' value.)
- *Systematic error* – a non-random error which is introduced by the way in which the experiment was conducted. This type of error gives a bias in the

result for the experiment, i.e. a consistent over- or under-prediction. Furthermore, this bias may not be eliminated or reduced by increasing the number of samples.[22] Using a ruler as an example, a random error would be the uncertainty from trying to interpolate a length that falls between two hash marks. A systematic error would occur if the measurement were made at a different temperature – because the wood or plastic from which the ruler is constructed will have a temperature coefficient of expansion.

Most experiments involve taking a number of measurements with more than one piece of apparatus. The errors associated with each measurement must be evaluated and combined in order to evaluate the accuracy of the result. One of the main topics underpinning this book is quantitative microscopy and therefore we leave it to the reader to ensure that their balances are zeroed, environmental factors have been considered and their Vernier scales are read correctly!

The previous sections in this chapter have described the removal of geometric and shading effects in images. These, combined with correct calibration of the microscope magnification, should ensure that the observed field of view is free from bias. However, this does not ensure that the experimental results are error free. The techniques used to measure features within the image and the selection of the image itself must also be free from bias. The subject of stereology addresses these aspects of experimentation and provides a 'toolbox' of mathematically proven techniques. Stereology covers a number of topics including uniformly random sampling techniques, counting rules and the inference of 3D features from 2D sections.

The need for rigorous sampling is quite apparent when considering the typical sample volume studied. For example, when analysing a small volume of translucent material, say $10\,\text{mm}^3$, a thorough experimental study might involve the inspection of 10 fields of view within 10 sections which are $5\,\mu$m thick. Using a typical microscope objective, the volume of material imaged is approximately one millionth of the total sample volume. Such a study surely must be conducted using a sound sampling method in order to avoid a gross bias in the results! Firstly, the physical sections must be chosen at random. With our example, this would involve randomly locating the first section in the range $x = 0 \ldots 1\,\text{mm}$, then sectioning at regular 1 mm intervals. Secondly, the images acquired with the microscope must be selected at random. Unfortunately, it is human nature to select, on the one hand, images that look more interesting, or images, which we think might give the results that are expected. Therefore, for our results to be valid, the field of view must be selected at random, however unsettling this may be.

Many experiments involve counting objects within the field of view in order to infer a number density or an estimate of the total number of objects within the sample volume. However, if we select a random section and location within a sample, a simple count of the number of objects within the field of view will

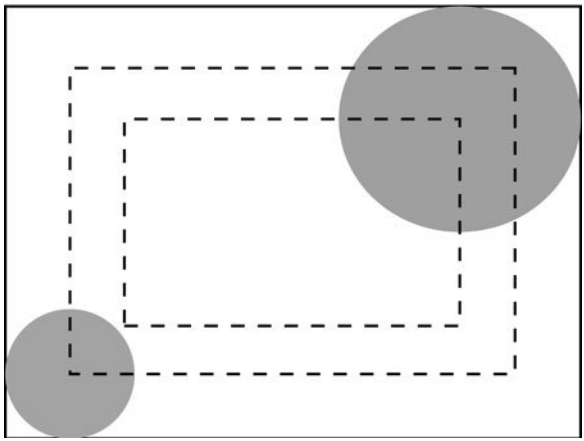


Figure 2.45 An illustration of the size bias introduced by edge effects. The centre of the smaller object, to be fully seen in the image, must be within the outer dotted rectangle. The centre of the larger object, to be fully seen, must be within the inner dotted rectangle.

give a biased result. The reason for this bias is illustrated in Fig. 2.45 where two objects of different size are seen within the field of view. In order to count an object it must be entirely within the field of view, therefore larger objects must be positioned further from the edge of the image. By comparing the shaded areas it can be seen that larger objects are less likely to be counted than smaller ones, i.e. our method is biased. We could calculate the effect of the bias on our results and compensate for it. However, a more desirable alternative would be to use a counting technique, which is itself unbiased. Figure 2.46 shows an unbiased counting frame, where objects are counted if they are within the counting frame, or intersect the acceptance line, but do not intersect the forbidden line. If there are no objects within the sample that are wider than the guard area, this counting method will give unbiased results regardless of object size, shape and density (see Gundersen[23]).

In the next two chapters, two optical microscopical techniques for materials science will be considered in detail. Both techniques generate vast quantities of image data because large sample areas (optical reflection microscopy) or large volumes (confocal laser scanning microscopy) are analysed and reconstructed. By application of the image processing scheme outlined in Fig. 2.28, the image data may be reduced significantly and the features of interest specified by appropriate parameters.

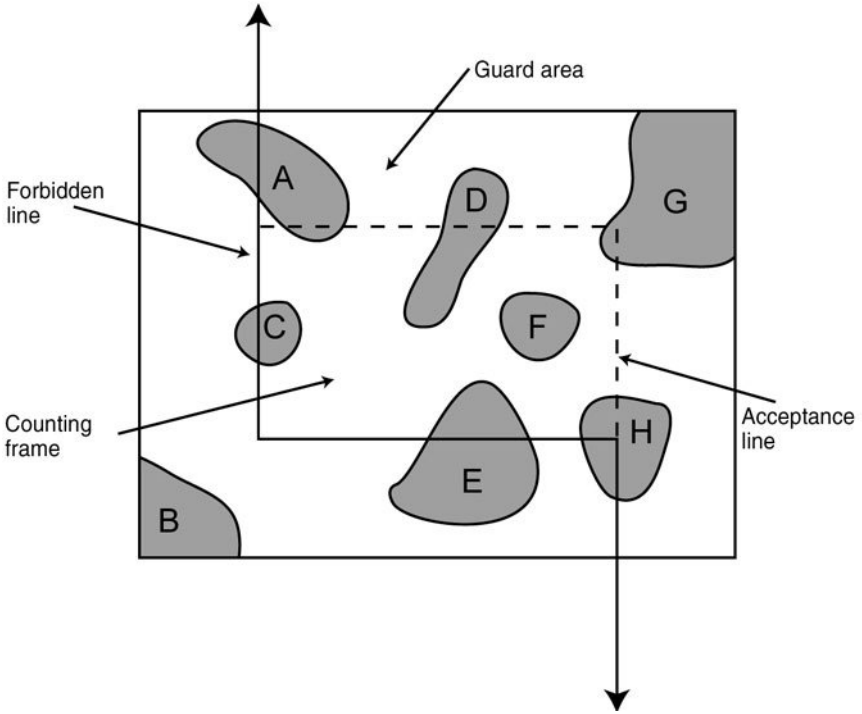


Figure 2.46 An unbiased counting frame for counting objects within 2D image fields. Objects of interest are only counted if they are within the counting frame, or intersect the (dotted) acceptance lines, but *do not intersect* the forbidden (solid) lines or the edge of the image. Therefore, in this example, only objects D and F should be counted.

2.10 References

1 Bolter J.D. (1984), *Turings Man: Western Culture in the Computer Age*, The North Carolina University Press.
2 Zadeh L.A. (1965), Fuzzy Sets, *Information & Control*, **8**, 338–352.
3 Shannon C.E. and Weaver W. (1949), *The Mathematical Theory of Communication*, Illinois University Press.
4 Bucci W. and Scrupski S. (1968) Special Report: The Transistor, *Electronics*, **41**(4), 71–130.
5 Kilby J.S. (1959), Semiconductor Solid Circuits, *Electronics*, **32**, 110–111.
6 Clarke A.R. (1983), *Vela Software Reference Book*, E.J. Arnold Publishers.
7 Johnson G.W. and Jennings R. (2001), *LabVIEW Graphical Programming*, McGraw-Hill.
8 Ifeacheor E.C. and Jervios B.W. (1997), *Digital Signal Processing: A practical approach*, Addison-Wesley.
9 Strum R.D. and Kirk D.E. (1989), *First Principles of Discrete Systems and Digital Signal Processing*, Addison-Wesley.

10 Moravec H. (1998), When will computer hardware match the human brain?, *Journal of Transhumanism*, **1**, March.
11 Berman G.P., Doolen G.D., Mainierie R. and Tsifrinovich V.I. (1998), *Introduction to Quantum Computers*, World Scientific.
12 Russ J.C. (1992), *Computer-Assisted Microscopy: The Measurement and Analysis of Images*, Plenum Press.
13 Huang J., Wang Y. and Wong E.K. (1997), *Check Image Compression: A Comparison of JPEG, Wavelet and Layered Coding Methods*, IEEE.
14 Murray J.D. and van Ryper W. (1994), *Encyclopædia of Graphic File Formats*, O'Reilly & Associates, Sebastopol, CA, USA.
15 Pennebaker W.B. and Mitchell J.L. (1992), *Jpeg: Still Image Data Compression Standard*, Chapman & Hall.
16 Gonzalez R.C. and Wintz P. (1987), *Digital Image Processing*, Addison-Wesley.
17 Press W.H., Teukolsky A.S., Vetterling W.T. and Flannery B.P. (1992), *Numerical Recipes in C*, Cambridge University Press.
18 Hough P.V.C., Methods and Means for Recognizing Complex Patterns, US Patent 3,069,654.
19 Xu B.G. and Yu L. (1997), Determining fiber orientation distribution in nonwovens with Hough transform techniques, *Text. Res. J.*, **67**(8), 563–571.
20 Ballard D.H. (1981), Generalizing the Hough transform to detect arbitrary shapes, *Pattern Recogn.*, **13**(2), 111–122.
21 Heisenberg W. (1925), Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen, *Z. Phys.*, **33**, 879–893.
22 Topping J. (1962), *Errors of Observation and their Treatment*, Science Paperbacks, Chapman & Hall.
23 Gundersen H.J.G. (1977), Notes on the estimation of the numerical density of arbitrary profiles: the edge effect, *J. Microsc-Oxford*, **111**, 219–223.

Part II

2D optical reflection and
confocal laser scanning microscopy

3

2D optical reflection microscopy

An expert is a man who has made all the mistakes, which can be made, in a very narrow field.

Niels Bohr (1885–1962) as quoted by Edward Teller, 10 October 1972 in *Dictionary of Scientific Quotations*, edited by A.L. Mackay

3.1 Introduction

The prime function of a traditional optical microscope is to present the eye with a faithful, magnified image of a 2D section plane and a typical research microscope is the Olympus BH2. In Fig. 3.1 the main components of the BH2 for recording images on 35 mm film, prior to the advent of computer-assisted microscopy with video camera image capture, are shown. The configuration of this particular microscope is for transmission studies through transparent films or thin sections, but the more common configuration for materials science would be the reflection mode, where the lamphouse would be located above the sample rather than below it as shown.

The traditional microscope was meant to be used qualitatively, in the sense that only in the late twentieth century did microscopists want to make quantitative analyses of the image field (instead of simply creating a freehand drawing of what they saw under the microscope). Then came the projection microscopes – enabling camera attachments, either 35 mm film cameras (as shown in Fig. 3.1) or large photographic plate attachments for the most sensitive applications and highest resolution studies. The problem with the film and photomicrograph approaches was that, during the developing process, artefacts could be introduced into the final images and also some form of digitisation (e.g. a digitiser pad) was necessary to take quantitative measurements.

Only in the past ten years has on-line image capture by video camera and the digitisation (and manipulation) of images been possible economically. The linking of the computer to the microscope opens up a raft of new possibilities, including high spatial resolution studies over large sample areas. A typical large area reconstruction is shown in Fig. 3.2. What cannot be appreciated from this

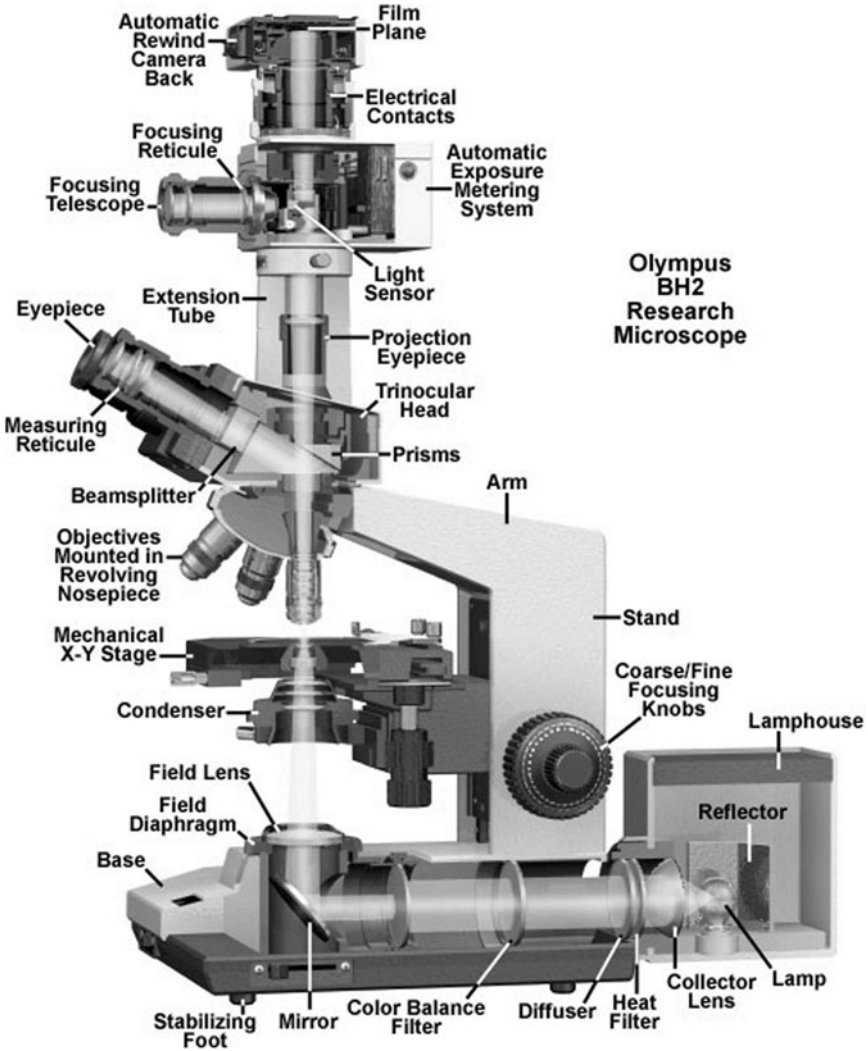


Figure 3.1 A typical research grade microscope – the Olympus BH2 – showing its operation in transmission mode. In this cut-away schematic diagram, a 35 mm camera attachment is shown for recording images.

large area image is the inherent quality of the spatial resolution, because the image is formed from hundreds of image frames like that shown in Fig. 3.3. As the authors' research team has been working on composite materials for the past twelve years, most of the comments and practical examples/images shown in this chapter relate specifically to fibre-reinforced composites. However, polymeric fibre-reinforced composites (with their different phases – matrix, fibres and, possibly, voids) are excellent examples of complex materials for illustrating general principles, which can then be extended to other specimens.

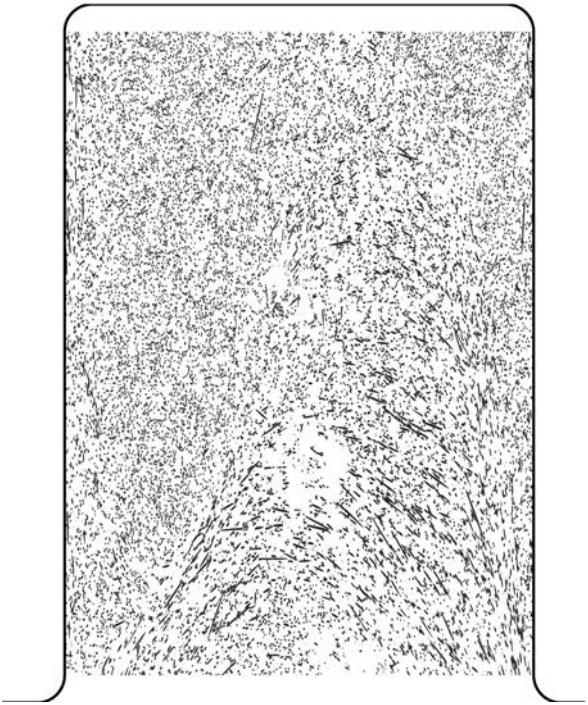


Figure 3.2 The primary function of a modern computer-assisted microscope system is to generate digital images of mesoscale size regions of the sample. This gear wheel tooth image area is 4 mm × 7 mm.

Composite materials are truly heterogeneous systems to characterise and model. The presence of stiff fibres embedded in the brittle matrix with different fibre packing fractions together with the possibility of voids makes for a finite element modeller's nightmare. The situation has not been helped by the fact that good quality 3D structural data has been lacking until a few years ago. Physical sectioning of composite samples and observation of the fibre images and void cross-sections under the optical microscope has been possible for the past twenty years. However, the tedium of manually collecting statistically meaningful, structural data has only recently been superseded by fully-automated, image analysis. In order to predict composite thermal and mechanical properties, one must have a reasonable, structural model and ensure that any simplifying assumptions are permissible. Also, in order to assess the effectiveness of the processing conditions for the manufacture of a complex part, one must be able to check the spatial distribution of fibre orientations (and voids) in the part. The ultimate goal must be to provide a sound theoretical linkage between the processing route and the final, global material properties. That goal will be achieved with good quality, 3D mesostructural data ('meso-' meaning on a length scale from a few microns through to centimetres, see Piggott[1]).

No matter what the application, it is clear that, in order to explore features within the field of view of an optical reflection microscope, the essential physical parameter is the reflection coefficient of the different phases that may exist within the field of view. If there is insufficient contrast between the images of interest and the general background, any post-processing of the digitised image field is likely to increase the errors of measurement. The most important consideration therefore is to find ways to improve the differences in reflection coefficient and this implies that sample surface preparation is of paramount importance.

3.1.1 Sample preparation

Although materials science technically covers all types of material, research into materials science usually entails dealing with hard samples rather than soft tissue type samples (as in the biosciences). If the microscopist is handling thin, soft materials (where the surface topology is not important), it is usual to place the sample between a slide and a glass cover slip. Clearly, in this case, one must use a set of objective lenses, which have been corrected for cover slip operation, as discussed in Chapter 1. However, if the sample is soft but thick, it could be microtomed into thin sections and these sections placed between slide and cover slip.

Most materials science samples – unlike biological specimens – are hard and only a small section (maybe a few centimetres by a few centimetres in size) will be studied under the microscope in one session. As it is assumed that the focal plane (or image plane) of the objective lens is flat, the sample has to be prepared so that it is easily handled and the surface (to be imaged) is horizontal to the sample support stage (or XY translational stage). If samples are being prepared for 2D optical reflection microscopy and analysis by image processing techniques, great care must be taken with both sectioning and the final polishing process. Physical sectioning of short fibre reinforcements, for example, can easily cause fibre pullout and the sectioning of aligned fibre composites can leave chipped fibres or fibre fragments at the surface, as shown in Fig. 3.3.

For cutting very hard materials, abrasive or diamond wheels are essential whereas diamond blade microtoming is appropriate for soft materials. The Struers ‘Rotopol-11’ machine used for polishing samples at Leeds is shown in Fig. 3.4. The first operation with a small specimen is to pot it in an epoxy resin block (see Fig. 3.35 later in the case study), so that it is easy to handle for polishing and sectioning. After potting, the rear surface of the resin block must be roughly polished (to define the reference surface on the support stage). As for the sample surface, polishing is normally undertaken with a set of graded silicon carbide (SiC) polishing mats starting out with the roughest (300 grade) and gradually introducing the finest (4000 grade). Recently, advice on composite sample preparation for image analysis has been published (Mlekusch *et al.*[2]) but

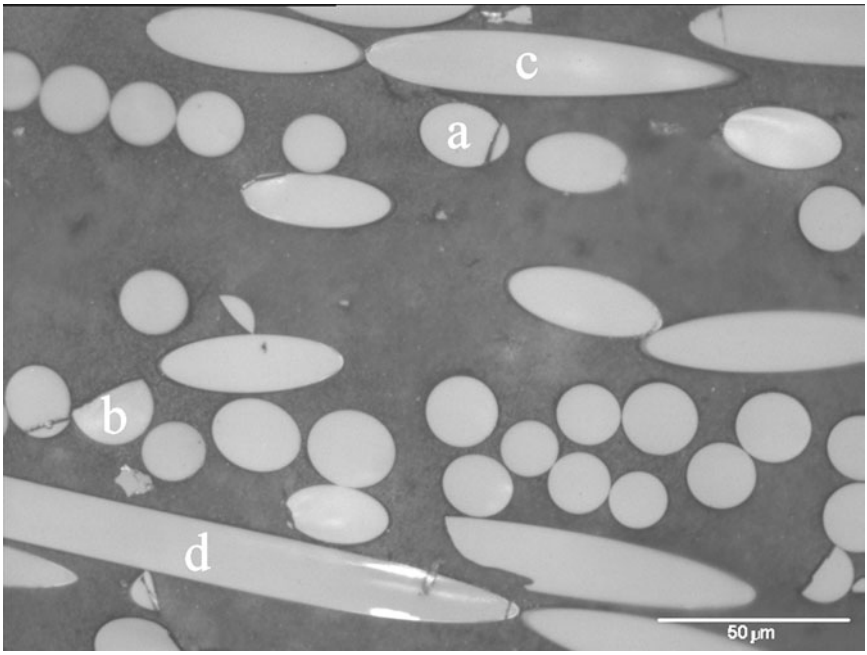


Figure 3.3 The large scale image in Fig. 3.2 was created from a collage of hundreds of higher spatial resolution images. This high-resolution frame (scale size 226 μm$\times$174 μm) shows typically high-resolution images of (a) split fibres, (b) fibre ends, (c) non-elliptical fibres (implying that they are all curved) and (d) fibres lying in-plane.

there are also many books e.g. Haynes[3] and Flynn and Powell,[4] that offer sound advice. A typical sample in epoxy resin is shown in Fig. 3.5.

Contrast issues for 2D optical reflection microscopy are no problem for carbon and kevlar fibres in polymer composites, because the reflection coefficient of the fibre phase is different from the matrix phase. However, glass fibres have very poor contrast with the polymer matrix because of the closely matched refractive indices and hence similar reflection coefficients. Therefore, special treatment of the surface of the polished sample is necessary to give better contrast to the glass fibres. In order to achieve the best quality contrast, one further step is used – oxidation of the matrix. Figure 3.6 shows the apparatus used at Leeds to perform the oxidation process.

The technique at Leeds for preparing composite materials has involved using an old sputtering rig and filling the chamber with oxygen (instead of a vacuum) so that the accelerated platinum ions from the anode collide with the oxygen atoms before hitting the specimen. In this way, the specimen is bombarded by oxygen ions, which oxidise the matrix rather than the reinforcing fibres. The roughened matrix surface gives low reflectance whereas the glass fibres have high reflectance, hence improving the contrast within the image field.



Figure 3.4 One of the polishing machines used in Leeds for sample preparation is the Struers Rotopol-11. Three samples, each potted in epoxy resin, may be polished at the same time.

Some samples may require chemical etching. An extensive list of suitable etchants for sample surface preparation can be found, for example, in Haynes.[3] Clearly, the glass fibre orientation characterisations that result from any 2D image analyser system attached to an optical microscope will depend crucially upon the quality of the sample surface.

3.1.2 Overview of the Olympus BH2 microscope

The microscope which has been used at Leeds for our computer-assisted image analyser system (explained in more detail in Section 3.2) is the economical Olympus BH2 microscope. A photograph of most elements of the Leeds system is shown in Fig. 3.7. Note the different position of the lamphouse to that shown in Fig. 3.1, which allows reflection mode operation. A more modern and suitable Olympus microscope nowadays would be the BX41F. However, the BH2 microscope has proved to be an efficient and reliable workhorse over the past

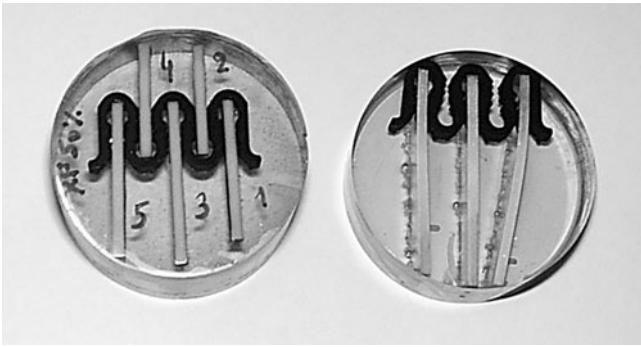


Figure 3.5 After sectioning, the samples are held by clips and potted in epoxy for ease of handling during the polishing process and also on the XY stage.

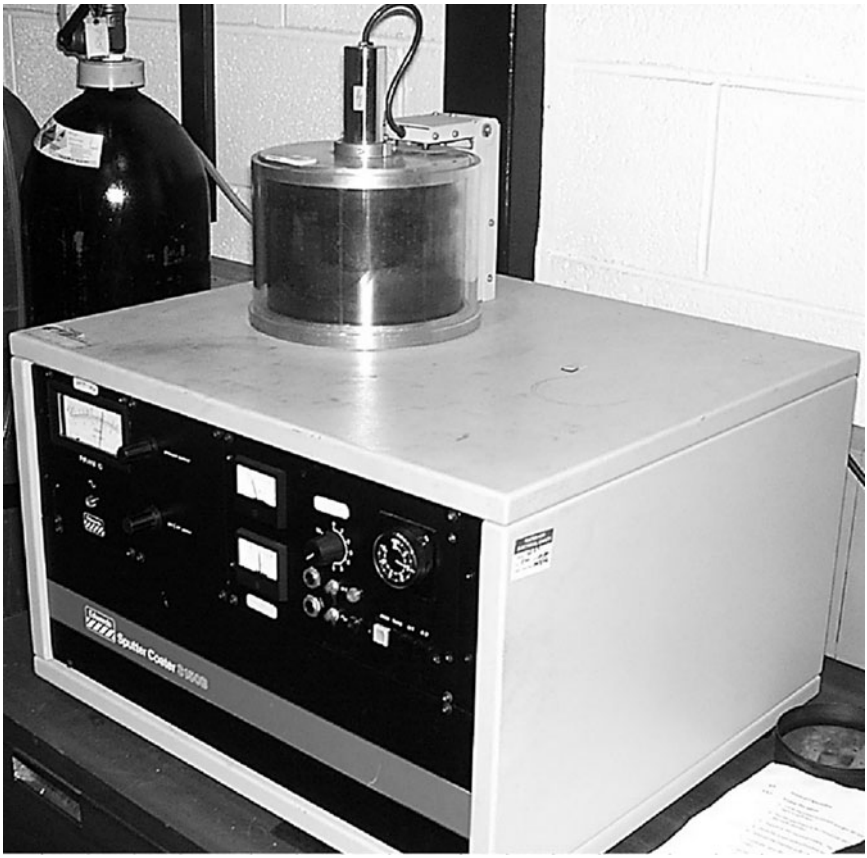


Figure 3.6 Photograph of a modified sputtering rig, which oxidises the matrix of glass fibre-reinforced composites (thereby ensuring that the glass fibre image cross-sections have greater contrast against the matrix).



Figure 3.7 Photograph of the Olympus BH2 microscope used in the Leeds 2D system, showing the Prior XY stage controlled by two stepper motors, the Z-drive unit and the Sony digital camera attached to the top of the microscope, via a MTV-3 adaptor.

twelve years. Clearly, for anyone starting with computer-assisted microscopy, any similar combination of basic microscope, XY stage and digital camera would be adequate for acquiring the images.

The light source has a manual, variable intensity control and a Sony XC-77 digital camera is C-mounted onto the microscope body. The basic set of lenses that has been used over the years for the fibre orientation and fibre length composite research consists of Olympus, infinity-corrected, objectives type MDPlan 5, 0.1 NA (IC5); MDPlan 10, 0.25 NA (IC10); MDPlan 20, 0.4 NA (IC20) and MDPlan 50, 0.75 NA (IC50). The IC number represents the magnification of the lens. The widest image field can be viewed by the operator through the binocular sidepiece, but a separate video monitor can ease feature location by presenting the operator with a real-time output from the video camera.

The Prior XY stage with integral stepper motors are shown together with the Prior automated Z-drive, which is necessary to incorporate for large area studies – because the flatness of the epoxy cannot be guaranteed to be within the depth of field of the microscope objective. Another useful tool is the sample rotation stage placed on top of the XY stage.

3.1.3 Practical considerations taking measurements on single image frames

With every new set of measurements, there is an art in acquiring and analysing the required data. For example, the most infuriating type of article/report discusses data without any comment on the errors associated with the measurements. There are many subtle ways in which errors can creep into the image data in computer-assisted microscopy, sometimes systematic errors and at other times random errors. By way of example, consider the apparently straightforward measurement of fibre orientations in fibre-reinforced composites. Although there is a specific case study on fibre orientations discussed in Section 3.3, it serves our purpose to discuss general issues surrounding these measurements here. In the fibre orientation measurement field, apart from Bay and Tucker,[5] Yurgartis,[6] Hine *et al.*,[7] Toll and Andersson,[8] Fischer and Eyerer,[9] Clarke *et al.*[10] and Mlekusch[11] errors were not discussed in any detail in composites research papers in the late 1980s and 1990s. This was a pity because, without an appreciation of the fundamental, instrumental limitations, it would be possible to use commercial software routines for fibre areas and fibre ellipticities and derive orientation distributions which are at best confusing and, at worst, totally inaccurate! This section is therefore a cautionary tale to highlight potential pitfalls, especially in the characterisation of fibre orientations from a sectioned plane.

Assuming that the fibres are smooth and circular in cross-section, a single 2D section through the sample would produce elliptical images, as shown schematically in Fig. 3.8. The elliptical image of the *i*th fibre yields estimates of fibre diameter, d_i, and the orientation angles ϕ_i and θ_i. A little thought shows that diameter, d_i, is equivalent to the minor axis, b_i, of the elliptical image, ϕ_i, is given by the orientation of the major axis, a_i of the elliptical image in the XY plane and the out-of-plane angle, θ_i, is given by equation [3.1].

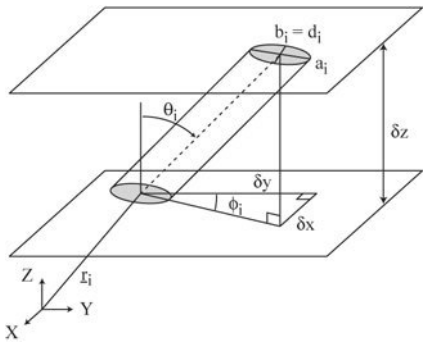


Figure 3.8 A schematic diagram showing a generalised fibre and the co-ordinate system adopted when discussing fibre orientations in our composites work. In order to classify a fibre in space, all of the parameters shown would have to be known, including the fibre length.

$$\theta_i = \arccos(b/a) \quad [3.1]$$

Note that ϕ_i is ambiguous because the same elliptical image would be produced by a fibre whose in-plane angle was ($180^o + \phi_i$) and see also Fig. 3.37. In many theoretical studies of material properties, this ambiguity may not be a serious limitation of the technique. The image processing methodology for deriving the fibre image ellipticities, and hence orientations, is discussed further in the case study in Section 3.3. However, there are a number of issues which should be considered before 'turning the handle' on image data using standard, commercial, software packages.

Firstly, one should consider illumination problems. The prerequisite for good quality images is that the illumination is uniform across the field of view. Clearly, the setting up of the condenser and light source(s) is crucial – so too the flatness of the sample within the image plane of the objective. During the development of the Leeds large area, 2D image analyser, discussed in the next section, an unexpected source of error in determining the fibre image area (and hence a potential candidate for fibre ellipticity error) became apparent. If the background illumination levels are systematically changing across the field of view, a constant pixel intensity threshold (used to determine those pixels, which are within a fibre image) will cause systematic orientation errors. Removing the sample and replacing it with a mirror surface will show if the illumination is indeed uniform. This is called 'flat-fielding'. The effect of an illumination gradient over the field of view and methods for its removal are discussed in Section 2.8.1.

Secondly, one should consider basic noise in pixellated images and how it might affect the measurement of the required characterisation parameter. As has been mentioned in Chapter 2, noise may be minimised by capturing a number of images of the same scene and averaging them. In this way, the random noise element partially cancels whilst the signal is reinforced. Once again, as an example, the effect of pixellated noise on the determination of fibre ellipticities, and hence fibre orientations, has been studied. Uncertainty in the best estimate of in-fibre pixels (not in-matrix pixels) will give a systematic error in the out-of-plane angle, θ_i, for fibre orientation determinations. As can be seen from Fig. 3.9, the square array of

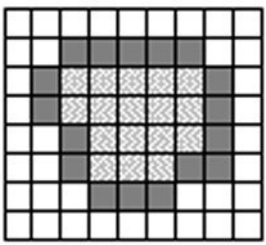


Figure 3.9 The digitisation process limits the precision with which a circular fibre cross-section can be established. As shown, depending on the magnification of the cross-section, it is difficult to distinguish a circular fibre from a nearly circular fibre when the pixels are square or rectangular.

pixel intensities cannot reproduce a perfectly circular image of a fibre cross-section. Clearly, with higher system magnification, the fibre image could occupy far more pixels within the image field (and hence give a better approximation to a circular image), but this higher magnification would be counter-productive, because far more frames would be needed to create a large area scan.

One should also check on the aspect ratio of the image field that is reproduced on the computer screen. The camera (if digital) must have square pixels on the active CCD chip if it is to digitise the field and produce a simple mapping into the array of pixel values. Some cameras have rectangular pixels, which mirror the aspect ratio of a typical computer screen. This strategy means that the scene on the screen shows the actual image field more faithfully because of the rectangular shape of the computer screen. However, if quantitative measurements are to be made from the digital data, the user must take care to correct for the fact that the true physical distance between *vertical* pixel centres may be different from the true distance between *horizontal* pixel centres! The effect on the determination of out-of-plane fibre angles of an error of five per cent in the aspect ratio is shown in Fig. 3.10. Clearly, if a known square grid pattern on a calibration slide is imaged, the appropriate aspect ratio can soon be discovered. Another interesting approach, which we have used in the past, is to analyse features within the same area on a sample. By comparing the results from an analysis in one sample orientation with the results when the sample is placed at 90º to the original orientation, the best value of aspect ratio may be deduced, as shown in Fig. 3.11.

Any commercial package used to determine the orientations of short fibres will most probably yield a frequency distribution of fibre orientations like the plot in Fig. 3.12. Even though this sample contained fibres which were oriented isotropically, it exhibits a peak around 15º and rapidly decreases towards an out-of-plane angle of zero degrees. The key issue here is that the apparent peak in

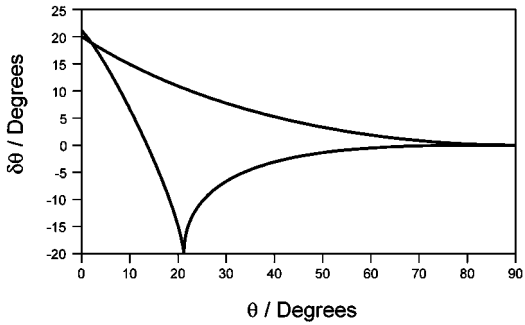


Figure 3.10 If the aspect ratio (the ratio of the height to width of each pixel) is incorrect by five per cent, there will be an error, $\delta\theta$ in the out-of-plane angle, θ, depending on the orientation of the fibre relative to the section plane orientation. The curves give the worst case scenario for these errors.

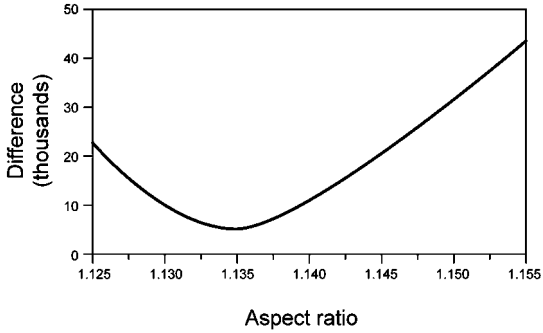


Figure 3.11 One technique for establishing the correct aspect ratio of the pixels is to perform a large area scan with the sample oriented in one direction and to compute the fibre orientation distribution of the thousands of fibres within the field of view. Then the sample is rotated by 90° and another large area scan is performed to generate another fibre orientation distribution. The difference between these distributions, for different values of the aspect ratio, yield a curve as shown and the minimum of the curve confirms the best value of the aspect ratio.

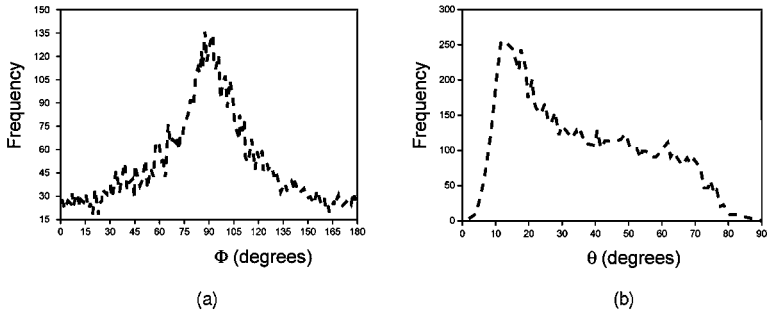


Figure 3.12 A typical set of fibre orientation distributions, for glass fibres in POM (polyoxymethylene), is shown where the raw orientation data have been presented. (a) The in-plane orientation distribution. (b) The out-of-plane angle distribution illustrates the peak around 15° to 20° and the decrease in apparent fibre numbers as the out-of-plane angle tends to zero and to 90°, even though the fibres were approximately isotropically distributed in this sample.

the orientation distribution around 10° to 15° is totally spurious. This spurious peak would appear even if the fibres were all highly aligned and oriented normal to the sectioned plane. There are three causes:

- the *mapping* of a circular image onto an array of square (or rectangular) pixels,
- the *solid angle effect* mentioned below, and shown diagrammatically in Fig. 3.13,
- the *height bias* introduced when the fibres are sampled by the section plane (as illustrated in Fig. 3.14).

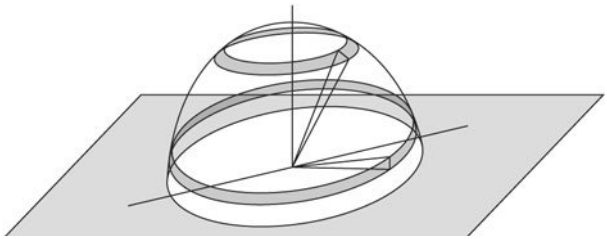


Figure 3.13 Another effect to consider, when deriving the fibre orientation distribution, is the solid angle effect. The solid angle for detection, $\delta\Omega$, of a fibre with out-of-plane angle between θ and $\theta + \delta\theta$ shown shaded) changes dramatically from nearly zero steradians (when the fibres are almost perpendicular to the section plane) to a maximum number of radians when the fibre is nearly parallel to the section plane.

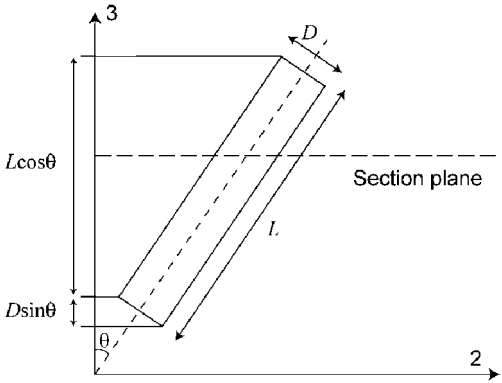


Figure 3.14 The probability of detecting a finite length fibre within the dataset when the specimen is sectioned depends on the angle of the fibre to that section plane and its length. Fibres that are perpendicular to the section plane have a greater chance of being detected.

Although it is obvious that a spurious pixel or pixels assigned to the true fibre image area could create the wrong fibre ellipticity value, it is quite difficult to estimate the error in fibre orientation from the 2D section plane data. However, in a recent article, the confocal laser scanning microscope technique (CLSM) was used to evaluate the actual errors created by any 2D measurements (see Eberhardt and Clarke[12] and also the CLSM technique discussed in Chapter 4).

The moral is that, contrary to a 'common-sense' view, the 2D sectioning technique will give better estimates of fibre orientation by analysing highly elliptical fibre images (rather than near-circular fibre images). Consider the problem of minimising orientation errors by large angle sectioning. Assuming circular cross-section fibres, and also reasonably uniformly directed fibres, the best technique for minimising the orientation errors is to section the sample at an angle of 45º, or even 60º, to the preferred fibre orientation (see for example Hine

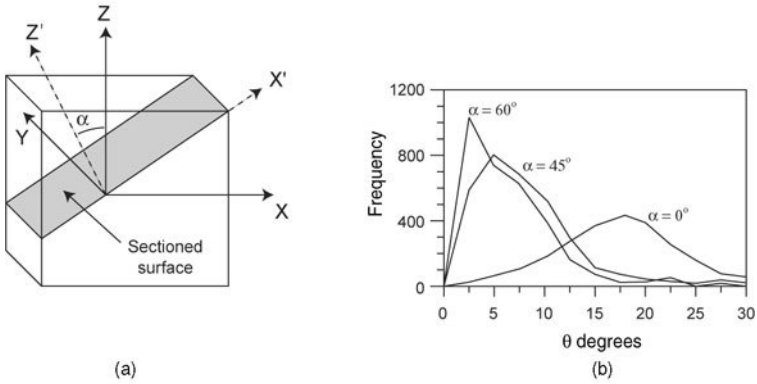


Figure 3.15 (a) One way to obtain better angular resolution, when the fibres are reasonably well aligned, is to section the sample at an angle, α, to the main fibre direction. In this way, instead of nearly circular cross-sections, the fibre cross-sections are elliptical and can be deduced with greater precision. All that is needed is to transform mathematically the fibre orientations back to the XYZ co-ordinate system. (b) The degree of alignment of the fibres may then be gauged more accurately as shown by the $\alpha = 60^{\circ}$ distribution.

et al.[13]). One can then mathematically transform back to that reference plane, which is perpendicular to the main fibre orientation direction. The appropriate transformation equations are given below:

$$\theta = \cos^{-1}(\sin\,\alpha\,.\,\sin\,\theta'.\,\sin\,\phi' + \cos\,\alpha\,.\,\cos\,\theta') \qquad [3.2]$$

$$\phi = \tan^{-1}\left(\frac{\cos\,\alpha\,\sin\,\theta'\,\sin\,\phi' - \sin\,\alpha\,\cos\,\theta'}{\sin\,\theta'\,\cos\,\phi'}\right) \qquad [3.3]$$

where α is the angle between the normal to the section plane and the preferred fibre direction, as shown in Fig. 3.15. Any spurious pixels have a smaller fractional error on the ellipticity value in this procedure. Clearly a practical problem arises if the fibres have a wide ($>45^{\circ}$) angular range of orientation values, or if the fibres are very short (see below). In this case, an alternative technical approach is needed such as the CLSM technique described in Chapter 4 or the X-ray microtomography technique described in Chapter 5.

Hence, we know for sure that the user of commercial software should be wary of the algorithm used for image ellipticity determination and, in general, it is certainly advisable to understand the basis of the algorithms that one wants to use in any commercial package.

Finally, one should consider the danger of making assumptions. One of the key assumptions made by all 2D image analysers of fibre images is that the fibre cross-sections are perfectly circular. If this is not the case, a little thought shows that the observed ellipticity of a fibre's image will be affected by any intrinsic

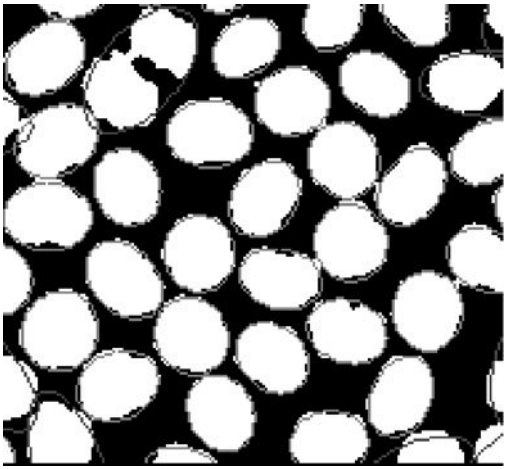


Figure 3.16 So far it has been assumed that the reinforcing fibres are smooth and have circular cross-sections. Some fibres, like these carbon fibres, have irregular, bean-shaped, cross-sections and hence great care has to be exercised in using the ellipticity technique for determining the fibre orientations.

ellipticity or irregularity in that fibre's cross-section. As an example of non-circular cross-section fibres, consider the high packing fraction, carbon fibre-reinforced sample shown in Fig. 3.16. Some carbon fibres have a 'fluted' shape around their circumference and others are kidney bean-shaped in cross-section like the ones shown! Clearly, by trying to fit a perfect ellipse to a blatantly non-elliptical object, a totally spurious orientation distribution will result. Yet another assumption often made, when researchers derive the fibre orientation, frequency distribution histograms, is that there is an equal probability of fibres being measured in out-of-plane orientation ranges, $\Delta\theta$. However, referring back to Fig. 3.13, it is clear that the probability of a fibre being detected within a certain solid angle, at a mean out-of-plane angle, θ, must be corrected for the area, ΔA, contained within that solid angle (i.e. $\Delta A = 2\pi r^2 \sin\theta\Delta\theta$). Hence, a meaningful frequency distribution of fibre orientations may be deduced. The probability of detecting fibres within a small angular range, $\Delta\theta$ about $\theta = 0^{\circ}$ is practically zero, because of the vanishingly small solid angle, but the probability of detection is high for θ values approaching 90º, as shown in Fig. 3.13. This is often referred to as the 'solid angle effect'.

The solid angle effect leads naturally to another issue – the application of stereological ideas or, inferring real 3D structure from arbitrary 2D sections. Before the 1970s, many researchers in materials science and the biological sciences did not take as much care as they could have and, hence, they came to conclusions about the 3D structure of their specimens based upon fundamentally biased data. As an example of the application of stereological principles, consider once again the frequency distribution of fibre out-of-plane orientations. Any

estimation of the frequency distribution of discrete fibre orientations from a single 2D section plane is a biased, height-weighted estimate of that distribution (see Gundersen *et al.*,[14] Cruz-Orive and Weibel[15] and Howard[16]). Height bias can be illustrated by considering the probability of a randomly placed, section plane intersecting each fibre. The probability depends on the projection of the fibre at right angles to the section plane, as shown in Fig. 3.14. If all of the fibres in the sample had the same discrete length, L, the probability that a fibre with out-of-plane angle, θ, intersects the sectioned surface would be proportional to cos θ. Hence the raw, 2D fibre orientation datasets from an actual specimen could be corrected (to some extent) for this effect by using the $\cos^{-1}$ weighting function.

Over the years, more complicated expressions have been suggested in the research literature.[5, 17] For example, to take into account the variation in fibre lengths and fibre aspect ratios (ratio of fibre length, l_i, to fibre diameter, d_i – not to be confused with the image pixel aspect ratio!), Möginger and Eyerer's proposed weighting function, $g(\beta, r, l_n)$, where $\beta = 90^{\circ} - \theta, r =$ mean fibre aspect ratio and $l_n =$ normalised fibre separation (dependent on the fibre volume fraction), is shown in equation [3.4].

$$g(\beta, r, l_n) = \frac{r}{\sqrt{r^2.\sin^4\beta + \left[\left(\frac{r+d}{l_n+d}\right)^2 + \left(\frac{r(l_n+d)}{r+d}\right)^2\right].\sin^2\beta.\cos^2\beta + \cos^4\beta}} \quad [3.4]$$

Bay and Tucker have also been considering the orientation states of short fibre-reinforced composites and the issue of fibres oriented within 20° of the section plane.[5] Assuming that the mean fibre length is L and the mean fibre diameter is d, there is a cut-off angle, given by $\theta_c = \cos^{-1}(d/L)$ below which the fibre images are elliptical in shape. (At angles $\theta > \theta_c$, the fibre image is non-elliptical because the section plane must cut through a fibre end.) Hence they suggested another, two-part weighting as shown in equations [3.5] and [3.6], where F_n is the weighting function and F_{90} is the weighting function at 90°.

$$\frac{F_n}{F_{90}} = \frac{1}{\frac{L}{d}.\cos\theta} \quad \text{for } \theta \leq \theta_c \quad [3.5]$$

$$\frac{F_n}{F_{90}} = 1 \quad \text{for } \theta > \theta_c \quad [3.6]$$

It is understood that more recent work by Tucker has further refined the weighting function above this critical angle, θ_c. In the case study (see Section 3.3.2), we discuss our alternative to these weighting functions.

The problem with all of these correction factors is two-fold. Firstly, they are 'model-based' (e.g. requiring assumptions to be made for the length distributions) and therefore it is impossible to evaluate the accuracy of the resulting fibre orientation distributions. As has been argued already, the best 2D data on fibre orientations are produced by large angle sectioning to preferred

fibre directions. However, if the sample consists of fibres with a wide range of orientations, this large angle sectioning and the transformation procedure breaks down and also, if a very high sectioning angle is used, any intrinsic curvature of the fibres will give a non-elliptical image. Secondly, the small errors introduced by not having a 'perfect' weighting function are insignificant compared to the fundamental errors introduced by the 2D sectioning and imaging methods (as discussed in detail by Eberhardt and Clarke[11]).

The worst case scenario is when the user suspects that the sample contains short fibres that are distributed isotropically. Clearly, one cannot use the technique of sectioning at a high angle to the preferred fibre orientation, because there is no preferred orientation! Note the lack of fibres in Fig. 3.12 which are close to $\theta = 0^{o}$ (due to the pixellation error and the solid angle effect). Also the lack of fibres at $\theta = 90^{o}$ (which is due to the fitting of ellipses to rectangular – or partial elliptical – images produced by short fibres lying in the section plane and the height bias effect). If two more 2D sections are taken at roughly orthogonal orientations to the first section, similar frequency distributions will result. The orientation data from these three plots would be consistent with an isotropic sample, but the data do not prove that the fibre orientations are isotropic, and the data do not allow for any measure of the deviation from isotropy. Once again, the CLSM technique can be used to test for isotropy, as described in Mattfeldt *et al.*,[18] if the sample is sufficiently transparent to allow for optical section planes to be placed within the material.

3.2 A large area, automated, high spatial resolution microscope (LAAM)

3.2.1 Introduction

It is undoubtedly true that a fast and efficient way to obtain large scale (cm $\times$ cm) image fields at high (sub-micron) resolution can help with the analysis of material mesostructure. Is there an alternative to using a computer-assisted microscope for the scanning of a large area sample and digitising the image? The answer is yes. If the sample is a very flat thin film or is a thin slice of material which has been produced by microtoming, reasonable spatial resolution images can now be derived using a conventional flat-bed scanner (for digitising documents) attached to the computer. A typical image, shown in Fig. 3.17, is of a dense, carbon fibre-reinforced thin film. Although the current devices cannot achieve the spatial resolutions of optical microscope systems, they are not too far away from competing. A typical system, at the time of writing, has a capability of scanning at 1200 dots per inch, which implies a spatial resolution of 20 microns. However, this resolution will no doubt approach optical microscopy resolutions in the future.

However, the low cost and versatility of microcomputers and image processing systems has enabled fibre orientation distributions to be derived

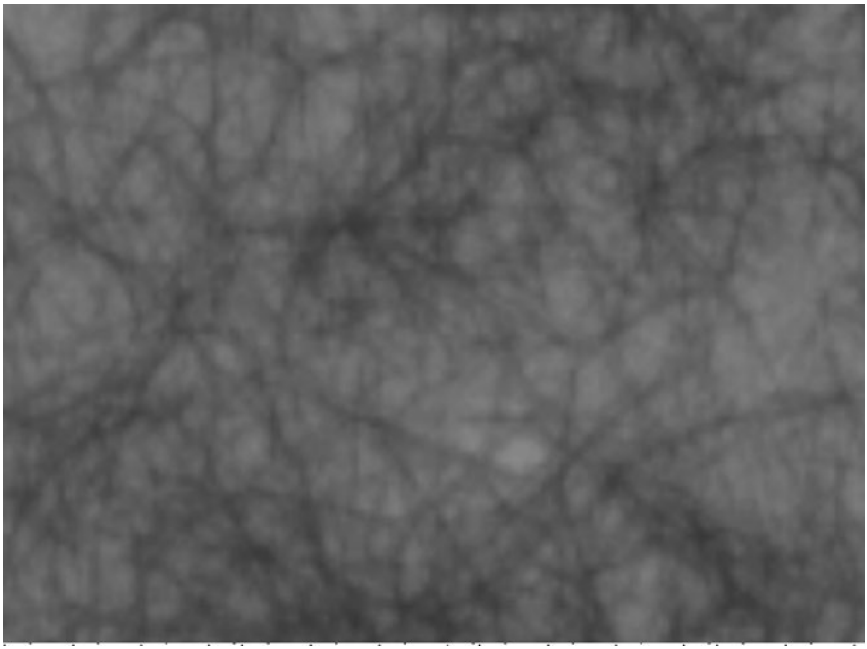


Figure 3.17 Large area digital images of thin specimens can be produced with the latest flat-bed scanners attached to computers. This image was created by a 1200 dots per inch scanner and represents a 6 mm×4 mm area of a carbon fibre-reinforced, thin film. Although the fibres cannot be individually distinguished, these spatial resolutions could determine clustering and other fibre morphological parameters.

from high spatial resolution measurements of hundreds (or, in some cases, many thousands) of fibre images on a polished, 2D section viewed with a conventional, optical reflection microscope. Recently, the current state of research into the use of image processing for the characterisation of polymer composites was reviewed by Guild and Summerscales.[19]

The Leeds 2D image analyser design was developed specifically for fibre orientation and fibre length research work. At the time of initiating the design and development, some researchers had used digitiser pads to take manual measurements of fibre images (e.g. Fakirov and Fakirova,[20] Yurgartis[6] and Toll and Andersson[8]). As the technology improved, image framegrabbers attached to economical microcomputers and running commercial image processing software, allowed more rapid measurements to be performed (e.g. Fischer and Eyerer[9] for fibre orientations in injection-moulded composites and Yurgartis *et al.*[21] for yarn shape and crimp angle in woven composites).

In the Leeds system, outlined in Fig. 3.18, a Sony XC-77DP CCD monochrome camera is C-mounted on a standard Olympus BH2 reflection mode microscope. A monitor displays the video signal from the camera for the benefit

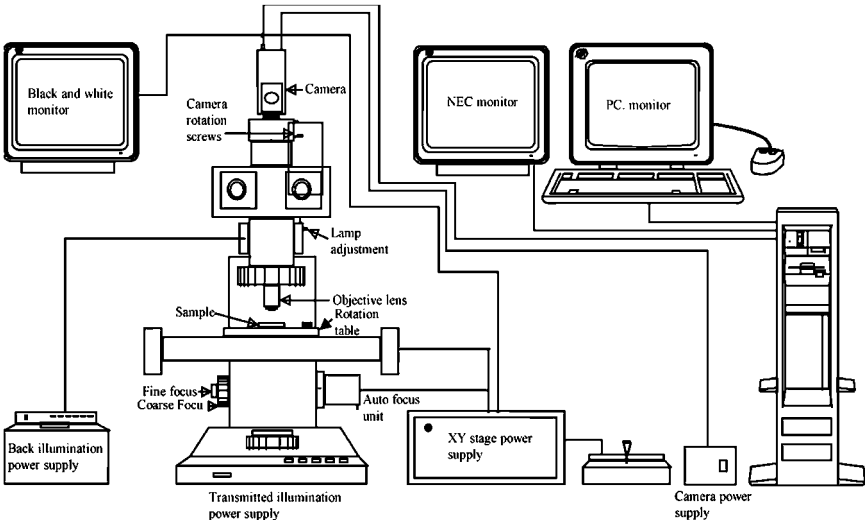


Figure 3.18 The hardware for the Leeds 2D system is shown in schematic form. The back-illumination power supply is for the reflection mode operation, which is the most common for materials science applications. The black and white monitor is useful to see, in real time, the image in the field of view. The other NEC monitors were used to keep the operator appraised of the fraction of the area scanned so far. However, in the latest version, the current image frame and previously analysed frames are shown on the main screen (see Fig. 3.26).

of the operator. The field of view displayed depends on the magnification of both the objective and eyepiece lenses. For fibre orientation work, an objective magnification of $20\times$ and an eyepiece magnification of $6.7\times$ gives approximately a $150\,\mu$m by $150\,\mu$m field of view for each image frame.

A fast Data Translation framegrabber type DT3155 fitted inside a standard, host microcomputer converts the XY image field into 512×512 pixels – at video rate frequencies. The pixellated images enable the XY field to be stored in the computer's video random access memory (VRAM) in readiness for digital algorithms to process the image frames.

A Prior XY stage is used to control the systematic scanning in X and Y over large sample areas. A Prior Z-drive through a microcomputer interface, incorporates an autofocussing algorithm (which selects that Z position giving the sharpest contrast along a line of the X–Y image). This feature is an essential item for large area studies, because of the limited depth of focus of the objective lens and the limits to the flatness of the sample surface. The monitor gives visual confirmation that the system is focussing each X–Y image frame and is scanning over the required surface area.

There are many commercial image analysers, e.g. Quantimet, Kontron on the market, and although they have excellent basic image processing facilities, they do not have the specialised features which are needed for high accuracy, fibre

orientation studies. The Leeds design has been developed over the past twelve years and the hardware/software has changed significantly over that period. Initially, the design was based on a small network of transputer chips and programmed in OCCAM but, with the demise of the transputer, the instrument had to be redesigned for the Intel Pentium chips and the software language rewritten from OCCAM into the C++ software language.

3.2.2 The initial transputer design

The transputer version of the Leeds 2D image analyser system has been described in detail elsewhere (see, for example, Davidson[22] and Clarke *et al.*[23, 24]) Basically, a network of four transputers, T800 from INMOS, were interconnected via their fast serial links (see Chapter 2) to form a small parallel processing network, as shown in Fig. 3.19. The system had a number of unique features for determining fibre orientations, which were retained in the later (Pentium-based) design:

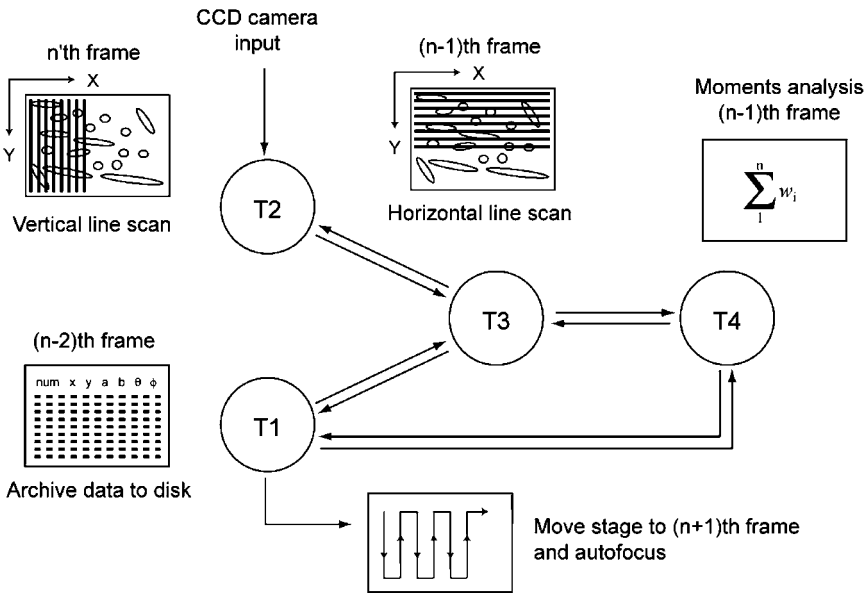


Figure 3.19 In the original Leeds design, a small network of four transputers was used to achieve speed-up of the fibre orientation process. The parallel function of each transputer is shown. As T2 performed the vertical pixel decoding of the *n*th image frame, the transputer T3 was decoding the horizontal pixels in the previous image frame. In this way, touching fibres could be split more easily. The fibre data from T2 and T3 was passed to T4 for calculation of the moments analysis of each fibre (so, in the example shown, it is working on data from the $(n-1)$th frame. Transputer T1 not only controlled the XY stage movement but also performed some necessary housekeeping like archiving earlier data to disk.

- autofocussing was carried out for each new XY image frame
- the image threshold was automatically set for each frame (to account for contrast changes)
- automatic fibre splitting of fibre images within the X–Y frame (see case study later)
- fitting of a partial image in one frame to the rest of that image in a neighbouring X–Y frame
- a 'fit factor' was assigned to each fibre image, measuring the quality of the elliptical fitting.

As all of the transputers were computing in parallel, the system was extremely fast in operation (considering that the clock speed of the T800 was only 20 MHz at that time). The hardware/software configuration was designed to process an image frame during the time taken for the X–Y stage to move from one position to the next. The system scanned sample areas measuring a few millimetres square with sub-micron resolution in X and Y (containing 20,000 fibre images) automatically in approximately 30 minutes. At the end of the scan, the images were automatically analysed for their ellipticities and fibre orientation distributions were obtained within seconds. Sadly, there were problems with the manufacture of transputers at higher clock speeds and these chips were phased out of production in the mid-1980s. Fortunately, Intel systematically increased the clock speeds of their chips and, at the time of writing, the Pentium chip has exceeded 2 GHz clock speeds. Hence, the attraction of a parallel processing network to achieve the required throughput of data has been superseded by a sufficiently fast serial microcomputer configuration.

3.2.3 Calibration of the image fields

In high spatial resolution, large area studies, an X–Y translation stage must be used to move the specimen under the fixed objective lens and hence effectively scan over large specimen areas by overlapping individual image frames. However, note that it will certainly be necessary to provide an auto-focus arrangement in the Z direction (the optical axis of the microscope). As the specimen is scanned, it is impossible to keep the specimen surface in focus (i.e. within the depth of focus of the objective lens) over many centimetres of travel in the X or Y directions. The X–Y translation stages (like the Prior system used at Leeds) are usually built from stepper motor drives but, more recently, piezoelectric drive stages have also become commercially available for large area scanning.

The general principle of the stepper motor is that a series of pulses is applied to the stator windings of the motor where each pulse causes the motor to rotate through a unit step displacement (see Fig. 3.20). The magnitude of the step depends on the design of the motor but is typically 1.8º, 2.5º, 3.75º, 7.5º, 15º or 30º. As the motor spindle (rotor) moves through a fixed step angle for each

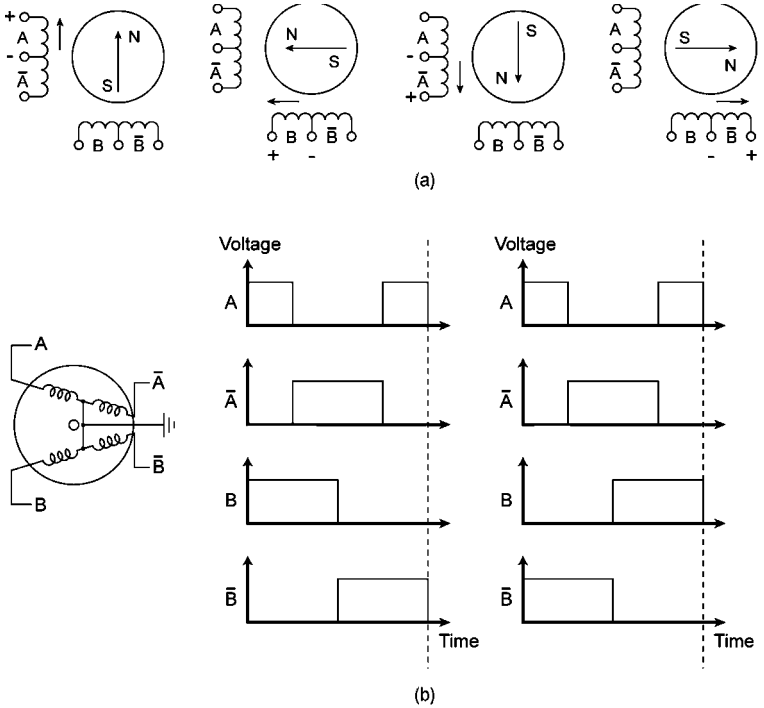


Figure 3.20 The operation of stepper motors is illustrated. Schematically, one can consider four windings driven by pulses and when the polarity of each drive coil is as shown, the rotor moves in discrete steps – forwards or backwards. For many stepper motors, the angle of rotation per series of step pulses may be 1 or 2 degrees. Finer rotation steps may be produced by appropriate external gearing or half-rotation steps (0.5° to 1°) by overlapping the drive signals.

pulse, it is often used in an open-loop control system and can provide highly repeatable rotational accuracy. Through the use of appropriate gearing and a lead screw, the rotational movement of the stepper motor spindle can be transformed into linear motion. Hence two stepper motors are usually employed in microscope XY translation stages to achieve repeatable, high precision movements. There are a number of different types of stepper motor, but the permanent magnet type offers the highest speed and highest efficiency.

Special stepper motor driver chips may be purchased which act as an interface between the serial, low power, pulse stream from a computer and the correct (high power) drive signals to the four windings on the stepper motor. A separate input control line to the driver chip determines the direction of rotation by altering the relative phasing of the four signals to the windings. The frequency of the incoming pulses determines the rate of rotation of the spindle, but note that too high a frequency would cause problems because of the finite time taken for the rotor to settle at its new orientation after each pulse.

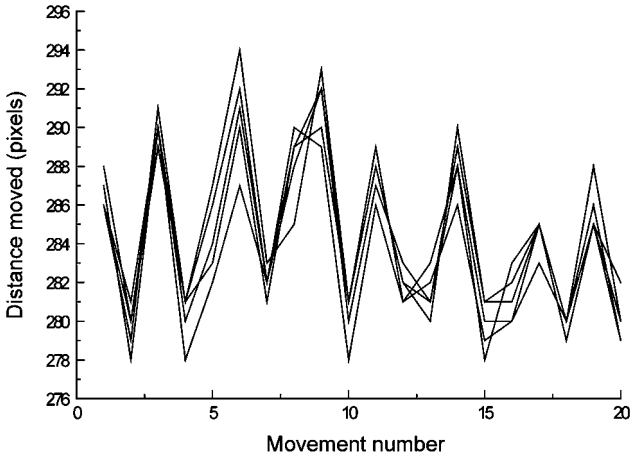


Figure 3.21 The rotary motion of the stepper motors is turned into linear motion via a lead screw and, if the lead screw is not perfectly circular (if it shows a slight ellipticity), a rotation by *n* steps will not produce the same amount of linear motion. By tracking images within the image fields, the number of pixels moved, when a fixed number of pulses are sent to the stepper, may be plotted. Note the repeatability of the pattern. The non-linear motion of the steppers can be corrected with a look-up table, provided that the area scanning starts from the same XY position.

When the computer sends a fixed number of pulses to a stepper motor (to determine the movement in X or Y), the stepper motor shaft rotation is directly related to the number of pulses. However, when converted into an X or Y linear movement, the actual linear movement produced is no longer proportional to the number of pulses, due to the eccentricity of the lead screw in the drive. This effect is only noticeable at the high magnifications employed and a typical plot of the variability is shown in Fig. 3.21. Hence, the manufacturers of X–Y stages prefer to quote the repeatability of positioning between two points (x_a, y_a) and (x_b, y_b) rather than absolute positioning accuracy from any initial starting point (x_i, y_i)! Note that the cyclic variations in the X or Y movement error are, to a first approximation, described by simple sine wave equations of the form $\delta X = a_1 \sin(b_1 X)$ and $\delta Y = a_2 \sin(b_2 Y)$ but over large area scans, there will be significant deviations from these simple equations. Therefore care must be exercised in removing this effect from the raw positional image data.

Fundamentally, there are three possible approaches to the calibration of the absolute positioning in X and Y over large scanned areas. The most expensive option is to back up the stepper motor movement with a laser interferometer to check on the absolute positioning of the moving parts! (This technique has been used on the large and powerful COSMOS facility at Edinburgh University for analysing large area, astronomical Schmidt plates.[25]) The second method (which is more economical) is to match some of the fibre images within one high spatial

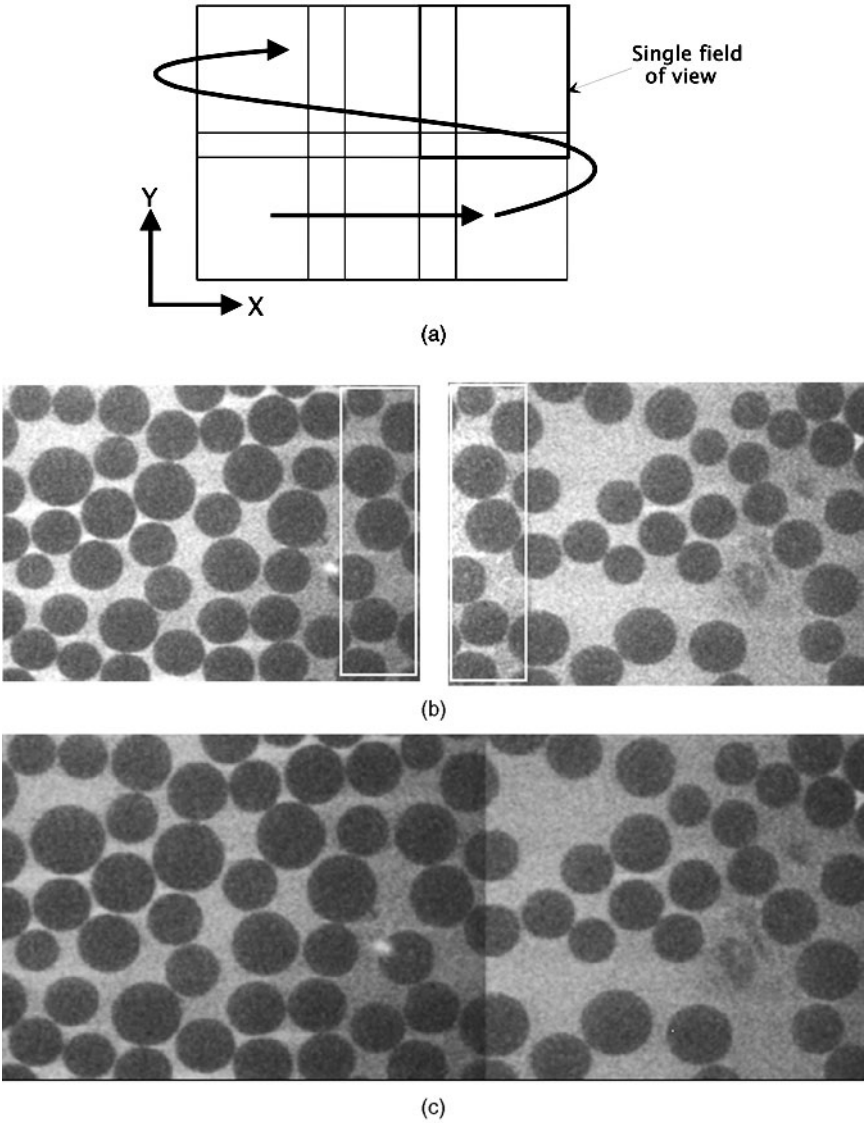


Figure 3.22 The technique for building up the raster pattern is illustrated with four image fields.

resolution XY frame with those same fibre images on the next (overlapped) XY frame (see Fig. 3.22). This technique only fails when there are no objects within the overlap region around the edge of the X–Y frame, which is an unlikely event during the analysis of most fibre-reinforced composites. (The variations due to the eccentricity in the lead screw plot, shown in Fig. 3.21, were derived using

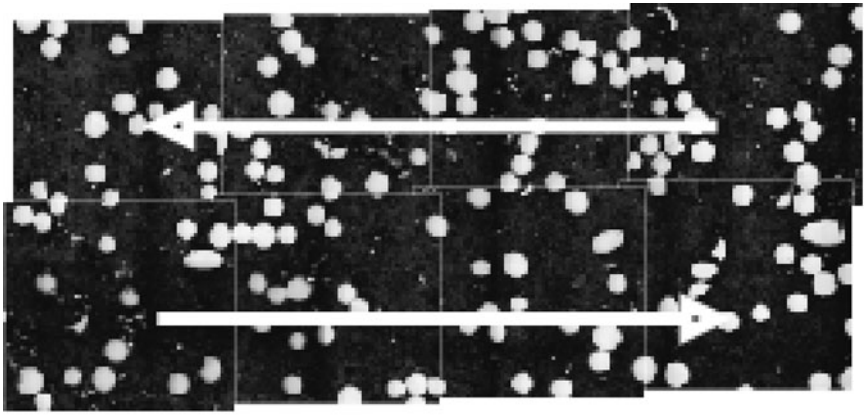


Figure 3.23 Larger area reconstructions use more sets of high spatial resolution images to form a collage – with overlapping frames. In practice, many hundreds of image frames may be used to cover a few square centimetres of sample area.

this technique.) Figure 3.23 illustrates how more frames gradually build up the large area scans.

However, there are occasions (e.g. scanning near to the edges of a complex part or when the sample has large resin-rich regions) when fibre images are lost from the field of view. In this case, a third method is to create a 'look-up table' of linear movements for a fixed set of pulses to the stepper motor. Note that this assumes that the large area scanning always starts from the same point in X and Y (and hence the XY stage always returns to that same point at the end of a large area scan). (If the XY stage were accidentally knocked, the operator would have to re-calibrate the stage movement and create a new look-up table.)

3.2.4 Parameters computed to characterise fibre orientations

The two angles (θ_i, ϕ_i) which determine the orientation of each fibre with respect to the X–Y section plane and the normal to this plane are computed using the 'moments' technique (see Stobie[25] and discussion in the case study, Section 3.3.3). This is not the only technique for determination of image ellipticities (see, for example, Wu & Wang[26] and the least squares fitting procedure of Mlekusch[11]). The moments technique has the merit that it is a 'single-pass' algorithm, which means that all partial sums required to compute the ellipticities are completed as soon as the image frame has been scanned pixel by pixel. The downside to the moments technique is that, if the fibre image is chipped or broken (i.e. non-elliptical), it has to be ignored otherwise an incorrect elliptical assignment will result, as shown in Fig. 3.24. In this regard, the Mlekusch least squares technique is more robust because these incomplete fibre images can be handled, as illustrated in Fig. 3.25. The relative errors introduced by the second

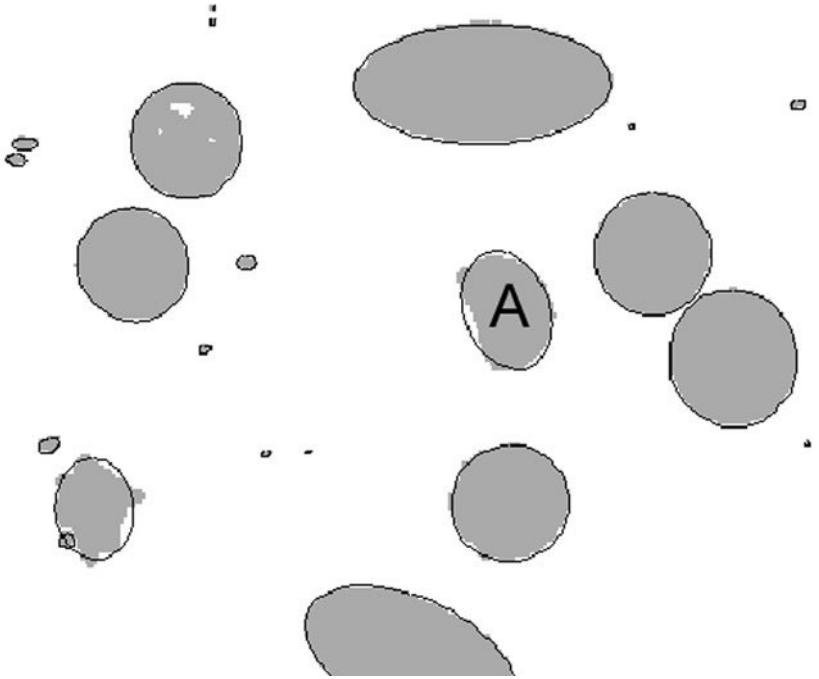


Figure 3.24 In this image frame, the elliptical cross-sections have been fitted using the second moments technique. The fitting procedure works well, but errors creep in when fitting an ellipse to fibre fragments like object 'A'.

moments technique and the least squares technique have been discussed in detail.[12]

Because of the possibility of chipped and broken fibres (and, in short fibre composites, fibre ends intersected by the section plane), it is vital to have a measure of the accuracy of the elliptical fit to the original fibre image data. A simple but useful parameter, the 'fit-factor', F_f, which evaluates the quality of the elliptical fit to the original pixel area, is defined later in Section 3.3.3. By plotting the semi-minor axis value, b, of the fitted ellipse against the fit factor, objects within the field of view could be identified depending on whether they are touching fibre images, incomplete or damaged fibre ends or good quality elliptical fibre cross-sections. Hence the system presents the operator with a histogram of the distribution of fibre diameters, after the required area has been scanned. Valid fibres are selected by defining a minimal and maximal acceptable minor axis value. These valid fibres are then displayed on the screen with the fitted ellipses overlaying the original, binarised, pixel image field.

A typical screen dump produced by the Leeds system during image data acquisition is shown in Fig. 3.26. The histogram plots at the bottom of the screen show the out-of-plane and in-plane orientation distributions for all fibres, which

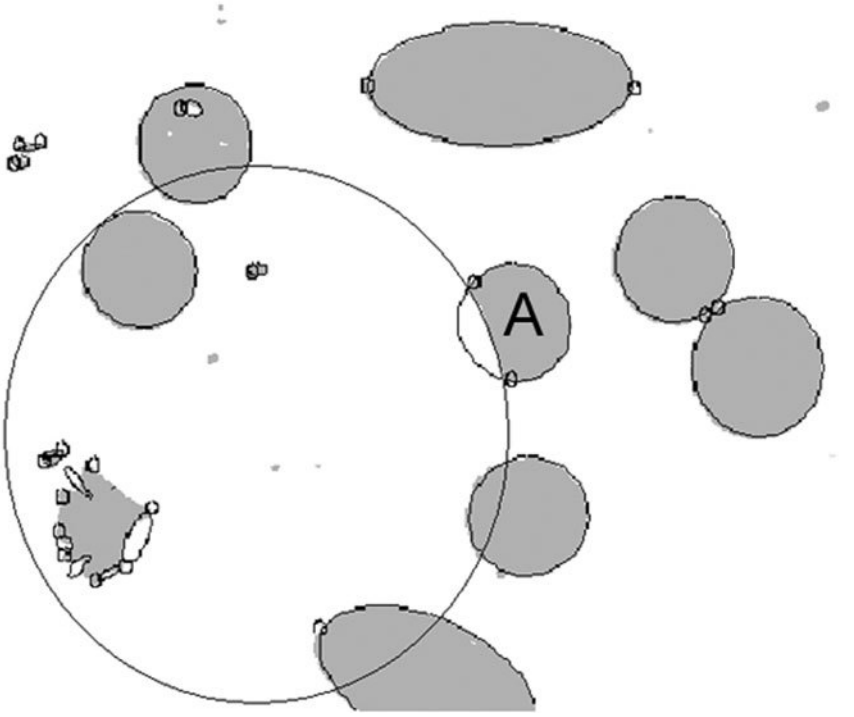


Figure 3.25 The same image frame as in Fig. 3.24 has been processed using the least squares technique for the elliptical fitting. Note the differences for fitting to fibre fragments, especially object 'A'. The second moments technique fits an ellipse to the entire object, whereas the least squares technique splits the perimeter and fits an ellipse to each section. The spurious ellipses that are too large or too small can easily be rejected by comparing their minor axis lengths to the known fibre diameters.

have been fitted. The large left-hand image frame shows the current image and the right-hand side image frame shows the previous binarised image frame with ellipses fitted.

The high spatial resolution, large area capability of the Leeds design enables the rapid scanning of any carbon, kevlar or glass fibre-reinforced composite to be carried out. The only downside to the 2D optical reflection technique is the ambiguity in the in-plane angle, ϕ, which is illustrated in the case study and Fig. 3.37. However, many of the classical ways of characterising fibre orientation distributions do not require this ambiguity to be resolved. Instead of a histogram of fibre orientations, many researchers looked for a single parameter which would be indicative of the overall fibre orientation state within the specimen.

The parameters that have been used to characterise fibre orientation states in composites were proposed to fit in with the state-of-the-art optical measure-

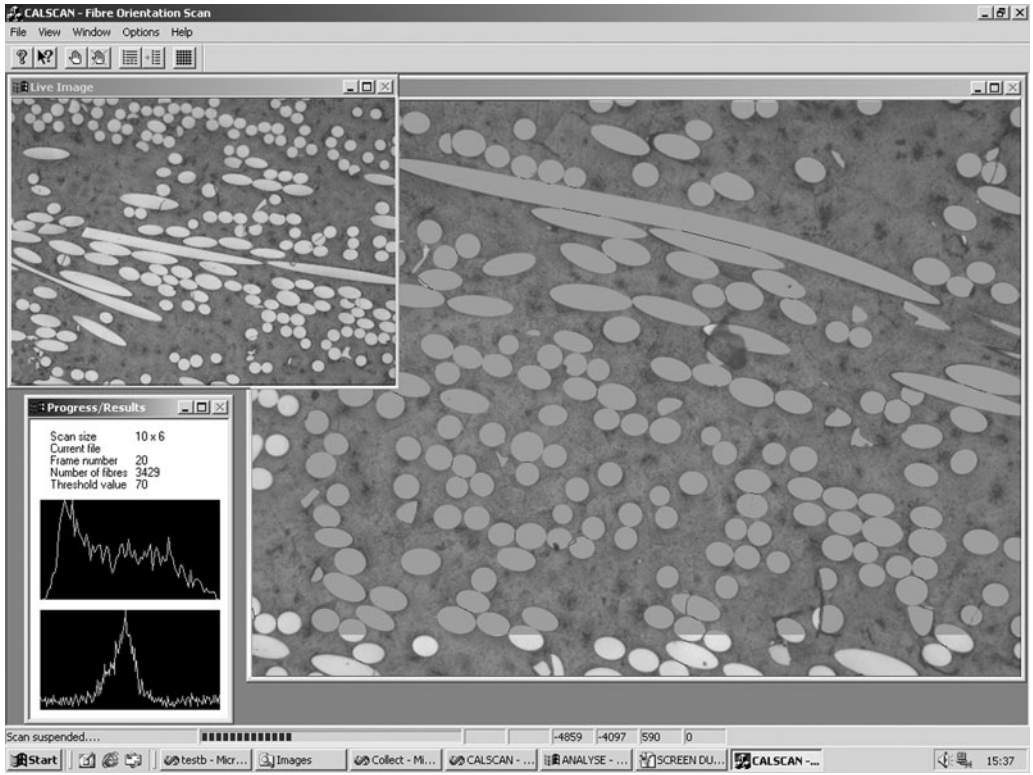


Figure 3.26 A typical screen dump during the XY scanning is shown. The smaller frame shows the current live image that is being captured and the larger frame shows the previous frame. The better contrast fibres, around the edge of this frame, are in the overlap region and do not have their ellipticities calculated on this particular scan. The top graph shows the out-of-plane angle (θ) frequency distribution and the bottom graph shows the in-plane angle (ϕ) frequency distribution of 3429 fibres, which have been collected within a few minutes.

ments that were possible at the time! Hence, as single 2D sections could be analysed by optical reflection microscopy, with or without automated image analysis systems, a number of geometrical coefficients were popular with experimenters, e.g. one of the first characterisations was Herman's orientation parameter, H_e, used by, for example, Fakirov and Fakirova[20] and O'Donnell and White[27]:

$$H_e = \frac{1}{2}\left(3\frac{\sum_i N(\theta_i).\cos^2\theta_i}{\sum_i N(\theta_i)} - 1\right) = \frac{1}{2}(3 < \cos^2\theta_i > -1) \qquad [3.7]$$

where θ_i is the angle of the fibre to the z axis and $< .. >$ denotes the average value. However, if the fibre orientations are not axi-symmetric about the Z-axis, this is an incomplete characterisation (because different orientation states could give the same value for H_e).

Fibre orientation states and their links with global elastic moduli have been studied extensively[28–30] using orientation functions, P_n, for example, $< P_4(\cos\theta) >$ which is related to the orientation averages $< \cos^2\theta >$ and $< \cos^4\theta >$ by the relationship

$$< P_4(\cos\theta) > = (35 < \cos^4\theta > -30 < \cos^2\theta > +3)/8 \qquad [3.8]$$

The derivation of tensor coefficients to represent fibre orientation states is also becoming popular[31] and the composites group at IKP, Stuttgart characterise orientation states in this way.[32] The general form of the second order orientation tensor is given below in equation [3.9] but see also the figures and associated discussion in Section 3.3.3. $\Psi(p)$ is the fibre probability distribution function and p is a vector parallel to the fibre. Higher order orientation tensors can be defined in a similar way.

$$a_{ij} = \int_p p_i p_j \Psi(p) dp \qquad [3.9]$$

The nine components of the second rank orientation tensor, a_{ij}, have a physical interpretation that is very similar to the components of the stress tensor. For example, if the fibres within the sample volume were isotropically oriented in 3D, the components would reduce to

$$a_{ij} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \qquad [3.10]$$

Apart from fibre orientations, researchers have also considered the characterisation of the spatial distribution of fibres within composites. Some

of these characterisations are mentioned here because the absolute locations of all fibre centres can be deduced over large areas using the Leeds system.

A Japanese group[33] have suggested a fractal approach to the characterisation of fillers (mainly particulates and short fibres) in composite materials. They introduced three micro-morphological parameters: ξ, the homogeneity distribution parameter; η, the anisotropy parameter of the first kind; and ϕ, the anisotropy parameter of the second kind. If A is the area of the observation window and d_p is the most frequent of the interparticle distances, d, and N is the total number of particles in the window:

$$\xi = \frac{d_p}{\sqrt{A/N}} \qquad [3.11]$$

$$\eta = \frac{\sum_i m_i \cos 2\varepsilon_i}{\sum_i m_i} \qquad [3.12]$$

$$\phi = \frac{\sum_i m_i \cos 2\varepsilon_i}{\sum_i m_i} \qquad [3.13]$$

Although the mathematical expressions for their anisotropy parameters of the first and second kind are identical, the definition of the angle, ε_i is different. Taya also argues that the degree of homogeneous distribution of fibres (or particulates) depends on the window size, L, and, if the number of fibres N within the window obeys fractal behaviour, $N(L) = \alpha L^{D_f}$ where D_f is the fractal dimension and α is a constant. Both α and D_f were proposed as potentially useful characterisation parameters.

In the case of continuous, well-aligned fibres, Yurgartis[34] has explored other characterisation parameters, e.g. 'included angles' (see Fig. 3.27) which are derived from nearest neighbour fibres, and Clarke *et al.*[35] have used the Leeds LAAM to investigate nearest neighbour in-plane orientation effects, because the moments technique over the large area scans gives the absolute central coordinates of each fibre (see Section 3.3.3). Pyrz[36] has also been concerned with the apparently random positioning of fibre centre coordinates seen on physical sections taken perpendicular to the preferred fibre directions. In common with a number of investigators, he has considered a Voronoi cell approach and various clustering parameters, like the second-order intensity function, *K(r)*, in his studies of material strength linked to composite microstructure. The function $K(r)$ characterises different fibre patterns in a 2D section and is defined by

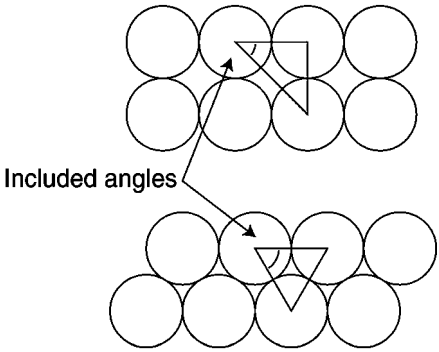


Figure 3.27 'Included angles' were used by Yurgatis to characterise the spatial structure of fibre-reinforced composites.

$$K(r) = \frac{A}{N^2} \sum_{k=1}^{N} w_k^{-1} I_k(r) \qquad [3.14]$$

In this formulation, N is the number of points in the observation area, A; $I_k(r)$ is the number of points within a circle of radius r (centred on one of the points), and w_k is a weighting factor to allow for image plane edge effects. In Fig. 3.28(a), the weighting factor for the circle of radius, r_1, is unity whereas, for circle radius, r_2, it is simply that fraction of the circumference contained within area A to the total circumference of the circle. Figure 3.28(b) shows plots of $K(r)$ for different types of spatial distributions from fibre clusters through to regular arrays of fibres. In practice, the real spatial distribution of fibres in a composite part is unlikely to follow any of these paradigms, but the $K(r)$ factor can indicate a tendency to cluster, rather than the distribution approximating to a random Poisson process.

Although the Leeds system gives the geometrical coefficients defined above, the most important parameter is the orientation tensor – as discussed further in the orientation case study, Section 3.3. The approach taken to characterise the spatial distribution of fibre orientations across a complex part is, firstly, to split the large area scanned into a manageable array of smaller areas (say $1\,\text{mm} \times 1\,\text{mm}$). Secondly, the appropriate orientation tensor coefficient is computed for each small area (see, for example, tensor coefficients shown in Fig. 3.44). This type of treatment readily identifies skin-core-skin regions in injection-moulded samples or distinctly different regions in woven fibre-reinforced composites (see Fig. 3.45 in the case study).

More complex types of characterisation can now be carried out. For example, 2D optical techniques have also been used to determine the waviness and/or misalignment of plies in multi-ply composite materials (see, for example, Camponeschi[37]). However, we believe that the best way to characterise fibre waviness (when the sample is sufficiently transparent) is by using the CLSM

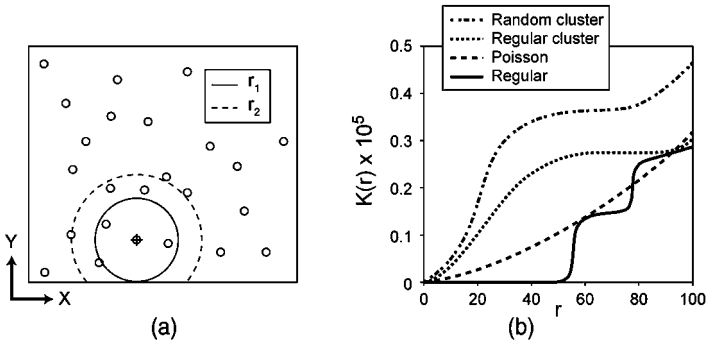


Figure 3.28 $K(r)$ function of Pyrz (a) shows the correction for edge of field and (b) is a plot of some particular types of fibre clustering on a 2D section plane, including the random Poisson distribution.

technique (see Chapter 4) or, alternatively, the X-ray microtomography technique (see Chapter 5).

3.2.5 Irregularly-shaped/odd fibres

So far, it has been assumed that the glass (or carbon) fibres reinforcing the composite samples were all of circular cross-section. Nowadays, composites research is being conducted with more exotic fibres, e.g. hollow glass fibres and also triangular glass fibres. The question is whether one can compute the fibre orientations for even these complex materials with a 2D image analyser system linked to an optical reflection microscope.

With hollow fibres, a simple flood-filling algorithm (as discussed in Section 2.8.6) must be used to pre-process the image fields before applying the moments technique (or another ellipticity algorithm). In Fig. 3.29(a) and (b), the raw image data together with the binarised image frame is presented and, in Fig. 3.29(c), the result of flood-filling, showing that the image data is ready for further processing, just like the conventional glass fibre-reinforced composites. Note that in the cases where the fibre images were incomplete, the flood fill has failed to block in the hollow centres.

However, the situation with non-circular cross-section fibres is more complicated. For example, a sample containing triangular cross-section glass fibres has been investigated by Leeds recently for the University of Bristol. The prepared sample surface contrast achieved is seen in Fig. 3.30. This shows that the pixel intensity histogram will be able to separate matrix, fibre and void regions. The images show an excellent three-colour segmentation: black for voids, grey for fibre and light grey for matrix areas. Samples were prepared by sectioning at 45º and 90º to the main fibre direction. The orientations of approximately 70 fibres were measured for each of these samples by manual identification of the nodes of each fibre's cross-section.

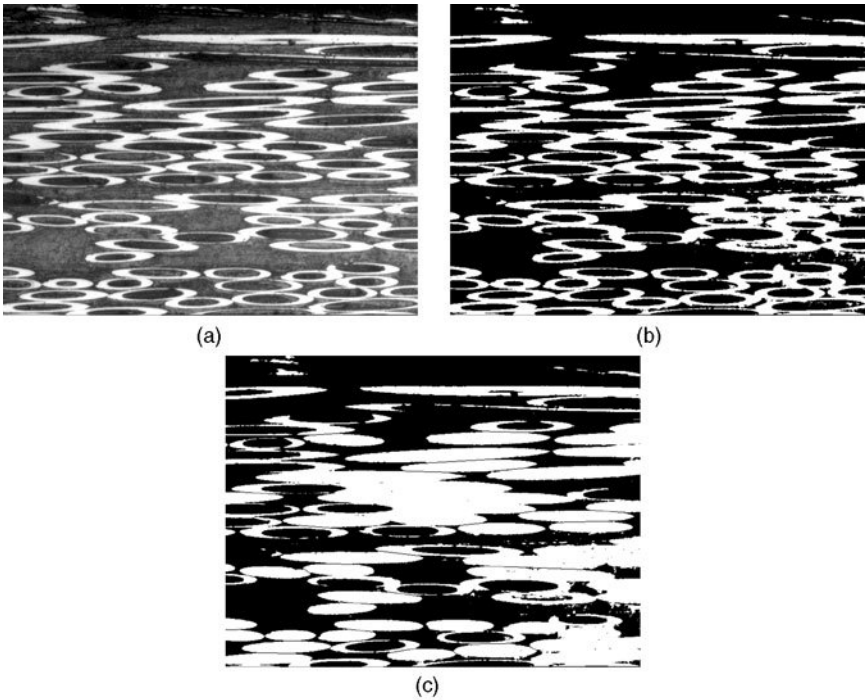


Figure 3.29 The effect of flood filling on hollow glass fibres: (a) the raw image data, (b) after binarisation and (c) after flood filling. Note that the flood filling has only been successful for fibres with complete perimeters after the binarisation process. Each image represents an area of 566 μm $\times$ 434 μm.

On the assumption that the fibre cross-sections are equilateral triangles, one can consider a fibre being sectioned at an angle, β, to the section plane and having a rotation angle, α, as shown in Fig. 3.31. The triangular fibre can be represented by three planes with the following normals:

$$\begin{aligned} A &= (\sin(\alpha), 0, \cos(\alpha)) \\ B &= (\sin(\alpha + \pi/3), 0, \cos(\alpha + \pi/3)) \\ C &= (\sin(\alpha + 2\pi/3), 0, \cos(\alpha + 2\pi/3)) \end{aligned} \quad [3.15]$$

The section plane is at an angle β to the Y axis of the coordinate system and is described by the equation:

$$E = (0, \sin(\beta), \cos(\beta)) \quad [3.16]$$

The intersections between these planes form the cross-section of the fibre. The line of intersection between two planes is orthogonal to the normal of each plane and is given by their vector cross product. Therefore the angle θ_{ab} is given by,

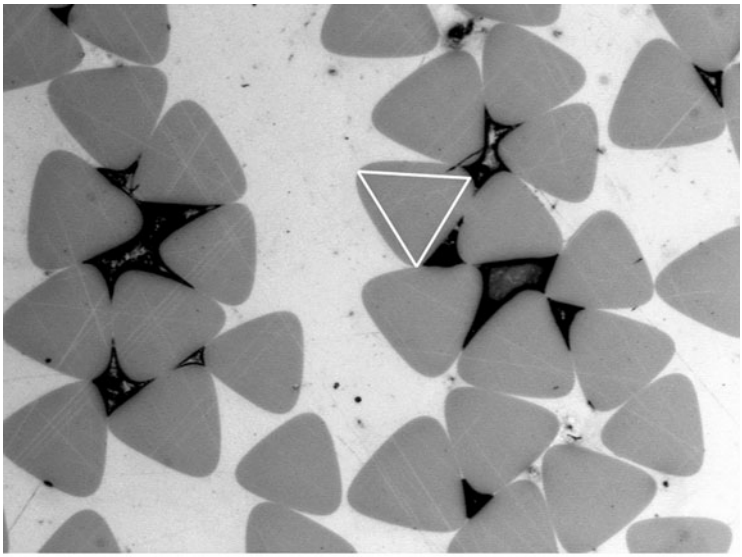


Figure 3.30 More unusual fibres – triangular shaped fibres (cut at 90 degrees to the main fibre orientation). Each image represents an area of 566 μm $\times$ 434 μm.

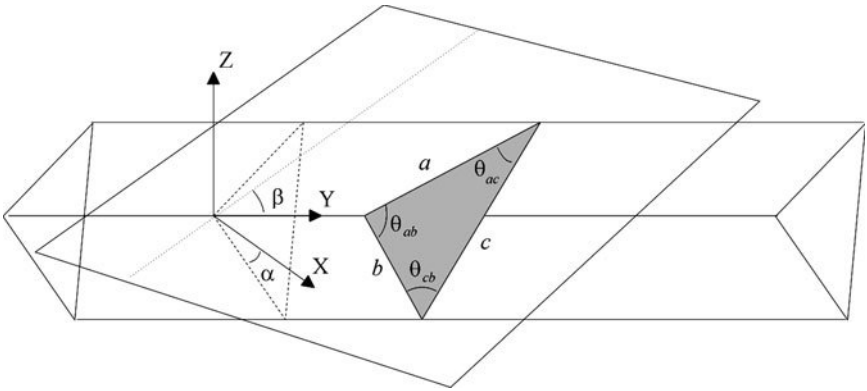


Figure 3.31 Schematic of a plane at a general set of angles to the coordinate system (α, β) intersecting a triangular fibre.

$$\cos(\theta_{ab}) = \frac{(B \wedge E) \cdot (A \wedge E)}{|(B \wedge E)(a \wedge E)|} \qquad [3.17]$$

which, after some manipulation, gives

$$\theta_{ab} = \pi - a\cos\left(\frac{\cos(\alpha + \pi/3)\cos(\alpha)[\cos^2(\beta) - 1] - \sin(\alpha + \pi/3)\sin(\alpha)}{\sqrt{1 - \cos^2(\alpha + \pi/3)\cos^2(\beta)}.\sqrt{1 - \cos^2(\alpha)\cos^2(\beta)}}\right) \qquad [3.18]$$

The relationship between θ_{ab}, α and β is shown in Fig. 3.32.

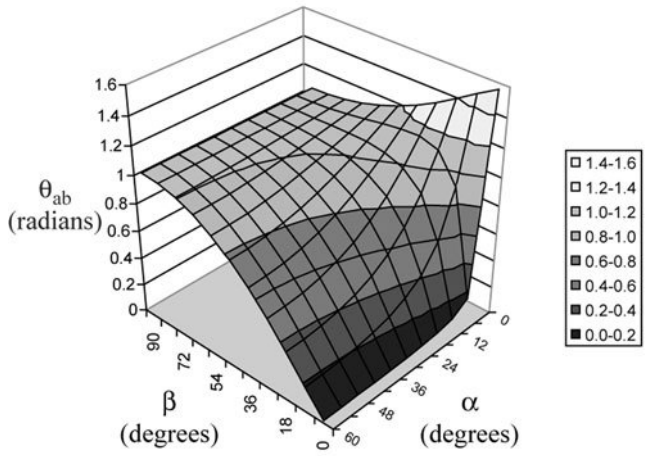


Figure 3.32 Relationship between θ_{ab} and a range of values of α, β for the triangular cross-section fibres. The units of θ_{ab} are in radians.

For any fibre cross-section as defined by $(\theta_{ab}, \theta_{ac}, \theta_{bc})$, it is possible to find the solution for the angle θ_{ab} on the surface of the plot in Fig. 3.32 and, similarly, for the other two angles, θ_{ac} and θ_{bc} giving:

$$(\alpha, -\alpha, \pi - 2\alpha/3, 2\alpha/3 - \pi, 2\alpha/3, -2\alpha/3) \text{ and } \beta \qquad [3.19]$$

For example, for the set of angular values ($\theta_{ab} = 0.521, \theta_{ac} = 1.994$ and $\theta_{bc} = 0.626$), the following are possible orientation states (α, β) for the fibre:

$$\beta = 21.6$$
$$\alpha = 34.38, -34.38, 85.68, 94.32, -25.56, 25.56$$

In order to assess the effect that this technique has on the fibre orientation measurements made on the 45º section plane, the equations can be re-worked for non-equilateral triangles, as shown in Fig. 3.33. The results, shown in Fig. 3.34, indicate a considerable spread of fibre orientations for the 90º section. However, inspection of the internal angles of the fibre cross-sections revealed that the original assumption made (that the fibre cross-sections are like equilateral triangles) is incorrect. The internal angles for three fibres are given in Table 3.1.

Table 3.1 Internal angles for three fibres

	θ_{ab}	θ_{bc}	θ_{ca}	Implied orientation
Fibre 1	63.1	60.8	56.0	66.6
Fibre 2	60.3	60.2	59.6	78.0
Fibre 3	62.5	60.6	58.1	73.5

Note: All measurements are accurate to ±1º.

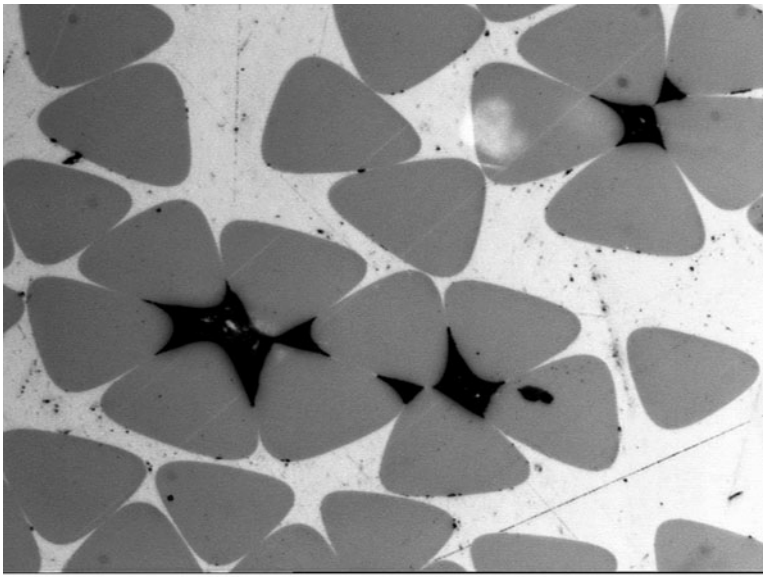


Figure 3.33 Triangular fibres sectioned at angle of 45° to main fibre orientation. Each image represents an area of 566 μm $\times$ 434 μm.

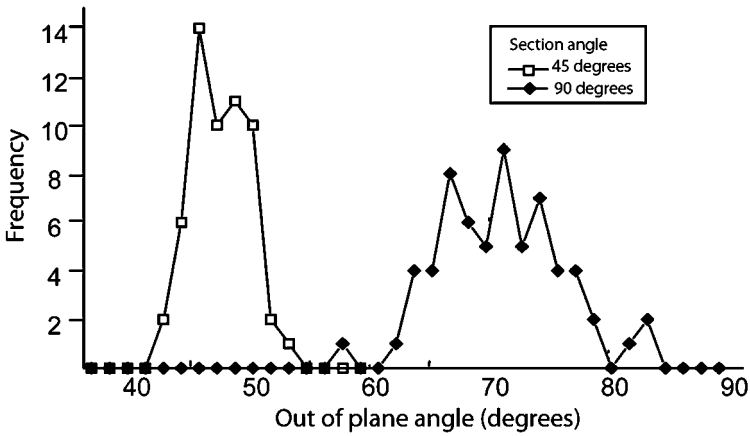


Figure 3.34 The histogram shows the calculated, triangular cross-section fibres out-of-plane, θ distribution for a section, which is cut normal to the preferred direction (90°) and, for comparison, when a section is cut at 45°. The tighter distribution at 45° indicates a better alignment of the fibres than is apparent from the 90° section.

3.3 Case study: fibre orientations within injection-moulded composites

3.3.1 Summary

This case study describes the measurement of fibre orientation within glass fibre-reinforced injection-moulded composite materials. The elliptical images of individual fibres are visible on prepared sections through composite materials and their orientations may be derived from the ellipticities. However, even small composite parts will contain millions of fibres, therefore automation is essential if realistic samples and accurate orientation statistics are to be obtained. In this case study, image processing routines for measuring fibre orientations are described and the stereological implications of deriving 3D characteristics from 2D sections are addressed.

3.3.2 Introduction

The most important feature of fibre-reinforced composites, their *raison d'être*, is the anisotropy in physical properties which they exhibit. The physical properties of fibre-reinforced composites are dependent on their composition (the matrix and fibre materials), and their structure (the physical arrangement of the fibres within the matrix). There are a number of structural properties which contribute towards the physical properties of the bulk composite, including the distribution of fibre lengths (see Section 3.4), void content, fibre orientation, waviness and curvature (see Section 4.7.6) and volume fraction. Of all these structural properties, fibre orientation plays the most significant role in dictating the desired anisotropy of the composites' physical properties (including tensile and compressive strength, elastic moduli and thermal conductivity).

Processing techniques for composite fabrication are designed to give control over the fibre orientation (and other structural properties) of the final component. The formulation of an optimal processing route can be a lengthy process involving experimental fine-tuning of the processing parameters by repeatedly manufacturing and comparing fibre orientations within a number of different samples. Recent advances in computer technology and simulations have made it possible to predict fibre orientation without the need for lengthy and expensive experimental work. However, computer simulation techniques have not yet removed the need for experimental fibre orientation measurements. On the contrary, an increasing number of research groups have become engaged in simulation work, and the need for accurate experimental data to validate their results has probably increased the demand!

The number of fibres within a composite component is typically many millions making it practically impossible to measure the orientations of every fibre in the sample. This problem can be reduced by the introduction of a small

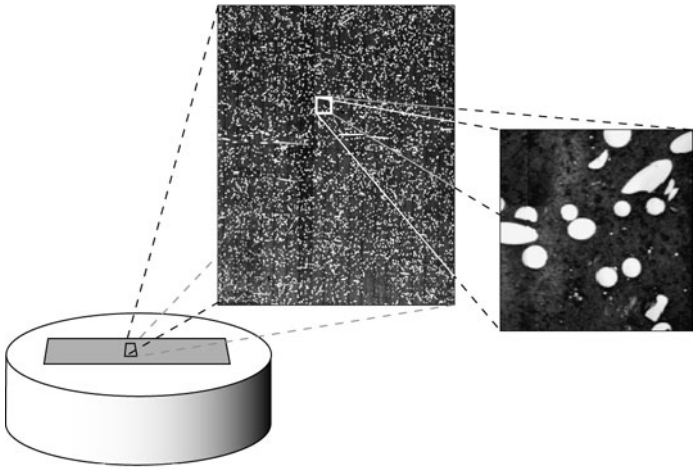


Figure 3.35 The specimen is potted in epoxy resin for ease of handling. The large area scan is built up of hundreds (or thousands) of high resolution image frames.

number of high-contrast tracer fibres, for example carbon tracer fibres can be used with optical imaging methods[38] (see Section 3.5.2) or leaded fibres in X-ray methods.[39] When the sample is imaged, the tracer fibres are clearly discernible from the surrounding matrix and standard fibres allowing their orientations to be derived. The orientation distribution of the tracer fibres is assumed to be a representative sample. However, fabrication of samples with tracer fibres is time consuming and expensive. Therefore techniques which image the composite without the need to prepare special materials are usually favoured.

Almost certainly the most popular instrument for determining the fibre orientation distribution within composite materials is the optical reflection microscope. Figure 3.35 shows a section taken from a glass fibre-reinforced composite, which has been prepared for analysis by a series of polishing steps. The contrast enhancement due to matrix oxidation ensures that the fibres are clearly visible by reflected light. If the fibres within the sample are cylindrical (as is the case for the majority of glass reinforcing fibres), the fibre cross-sections appear as ellipses whose degree of ellipticity is determined by their orientation with respect to the section plane, as illustrated in Fig. 3.36. It is possible to obtain a rough overview of fibre orientation distribution by eye from low magnification images, with gross differences in fibre orientations being visible. However, for quantitative comparisons of fibre orientation, further processing, at a higher magnification, is required.

An ellipse can be characterised unambiguously by its centre position (x_c, y_c), major and minor axis lengths, a and b respectively, and its in-plane orientation

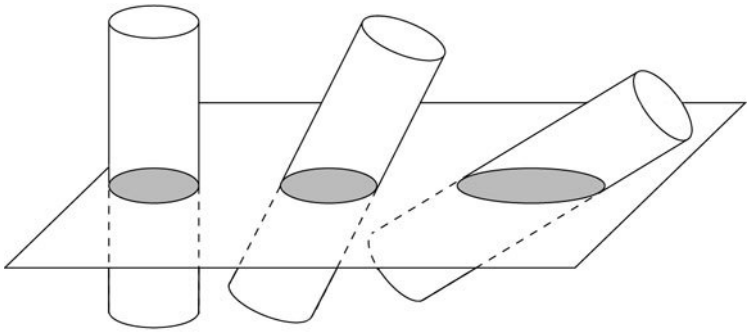


Figure 3.36 The ellipticity of a fibre cross-section depends on its orientation with respect to the section plane.

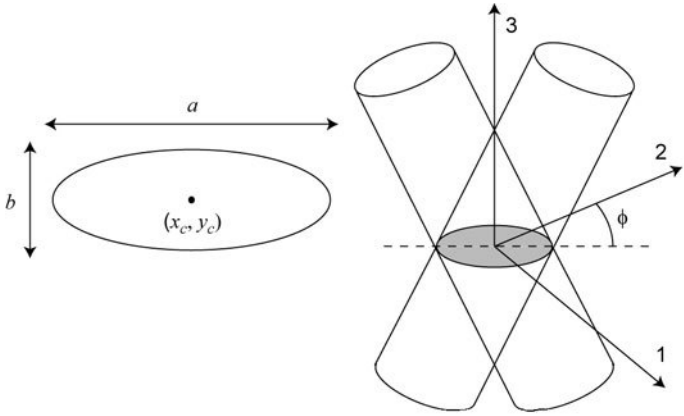


Figure 3.37 There is a fundamental ambiguity in the determination of the in-plane orientation of the fibre, but in many cases, this effect is unimportant.

angle ϕ, as illustrated in Fig. 3.37. The orientation of a single fibre can be expressed in polar coordinates by the two angles (θ_i, ϕ_i) and these can be determined from an elliptical cross-section as follows:

$$\theta_i = \arccos(b/a)$$
$$\phi_i = \phi \qquad [3.20]$$

The two possible values for ϕ_i are the result of an ambiguity in determining the orientation of a fibre, due to fibres with orientations ϕ and $\phi + 180$ having identical cross-sections. Therefore, the problem of measuring fibre orientation reduces to the problem of measuring the axis lengths and orientation of the elliptical fibre cross-section.

Some of the earliest fibre orientation measurements were performed by hand when photomicrographs were acquired and the axis lengths were determined

using verniers.[20] Clearly, this labour-intensive method was only able to acquire a relatively small set of fibres because of the time taken to analyse each fibre and the limits to the patience of the researcher. A slight improvement to the manual measurement technique was achieved with the introduction of a digitiser tablet.[5] With this interactive method, the researcher located the four endpoints of the ellipse axes and the computer calculated the axis lengths. More recently the increase in affordable computer power has fuelled research into image analysis, which automatically extracts the orientation data from these images.[9, 11, 22, 40, 41]

The identification and characterisation of elliptical objects is common in the field of machine vision, with any circular object appearing elliptical if viewed at an angle. A popular method, which is widely employed (because of its versatility and robust nature) is the Hough transform. As demonstrated in Section 2.8.5, the Hough transform is adept at locating partial or noisy features. However, it is relatively computationally intensive, involving the plotting of each point from the image plane as a curve in the parameter space (and an ellipse requires a five-dimensional parameter space).[42] Fibre diameters are normally found to be within a small range of values which reduces the size of the parameter space. This method has been employed successfully to locate fibres within reflected light micrographs.[43]

The oxidation stage of sample preparation ensures excellent contrast between fibre and matrix. Therefore, a straightforward binary threshold can achieve segmentation of the image into regions of fibre and matrix (see Section 2.8). Binary image processing techniques often provide a speed advantage when compared to corresponding methods working with the full greyscale image (such as the Hough transform). Therefore, binary processing methods are a good choice for this application where fast algorithms are required to analyse the large number of fibre cross-sections that are typically present. A method that was originally used for the analysis of astronomical data,[25] has been adapted and applied to characterise these elliptical cross-sections.

3.3.3 An automated fibre orientation measurement design

The reflected light microscope images are first segmented into regions of fibre and matrix by the application of a threshold. The excellent contrast between fibre and matrix leads to a highly pronounced bi-modal intensity histogram (see Section 2.8), making it a straightforward task to identify the optimum threshold intensity value. The binary images are further segmented into individual fibre cross-sections by region growing (as described in Section 2.8.6). Pixels within a fibre image have an assigned value ‘1’, others are assigned ‘0’. During the pixel-connectivity routine, the following sums are computed and stored for each object:

$$
\begin{aligned}
S_N &= n \\
S_X &= \sum x_i \\
S_Y &= \sum y_i \\
S_{XX} &= \sum x_i^2 \\
S_{YY} &= \sum y_I^2 \\
S_{XY} &= \sum x_i y_i
\end{aligned}
\qquad [3.21]
$$

where x_i, y_i is the location of the current pixel in the connectivity routine (with respect to an arbitrarily defined origin), and n is the total number of pixels in the fibre cross-section. The first moments of each cross-section are calculated as follows:

$$
\begin{aligned}
x_c &= S_X / S_N \\
y_c &= S_Y / S_N
\end{aligned}
\qquad [3.22]
$$

giving the centre location x_c, y_c of the fibre cross-section. The net second moments are calculated from the following equations:

$$
\begin{aligned}
M_{XX} &= \frac{S_{XX}}{S_N} - \frac{S_X^2}{S_N^2} \\
M_{YY} &= \frac{S_{YY}}{S_N} - \frac{S_Y^2}{S_N^2} \\
M_{XY} &= \frac{S_{XY}}{S_N} - \frac{S_X S_Y}{S_N^2}
\end{aligned}
\qquad [3.23]
$$

The second moments can be used to determine an equivalent ellipse (i.e. an ellipse with the same zeroth, first and second moments),

$$
\begin{aligned}
a^2 &= 2(M_{XX} + M_{YY}) + 2\sqrt{(M_{XX} - M_{YY})^2 + 4M_{XY}^2} \\
b^2 &= 2(M_{XX} + M_{YY}) - 2\sqrt{(M_{XX} - M_{YY})^2 + 4M_{XY}^2} \\
2\phi &= \tan^{-1}\left(\frac{2M_{XY}}{M_{XX} - M_{YY}}\right)
\end{aligned}
\qquad [3.24]
$$

The orientation of the respective fibre is then calculated from equation [3.20].

The second moments technique described above relies on the fibres having circular cross-sections and for the majority of glass fibre reinforcements this is usually the case. Experimental measurements using electron microscopy have determined that the upper limit to the intrinsic ellipticity is $e = 0.011$ (e is computed from the ratio of axis lengths, $e = 1 - b/a$).[44]

Because of the possibility of chipped and broken fibre cross-sections (and in short fibre composites, fibre ends intersecting the section plane) the features

within the image must be classified in order to remove these unwanted objects. The minor-axis length of a fibre cross-section is equal to the fibre diameter. Therefore, if the upper and lower limits of fibre diameter are known, this parameter can be used to remove some of the polishing artefacts. Another simple parameter, the fit factor *f*, which evaluates the quality of the elliptical fit is defined by,

$$f = \frac{\pi ab - n}{n} \qquad [3.25]$$

where πab is the area of the fitted ellipse and n is the actual measured pixel area. Clearly, a perfectly elliptical fibre cross-section will have the same area as its fitted ellipse, hence $f = 0$. Fibre cross-sections that are slightly irregular due to the presence of dirt or scratches have a fit factor of, typically, $0 < f < 0.1$, whereas fibre fragments will be highly non-elliptical with a fit factor in the range $1 > f > 0.1$.

The decision lines illustrated in Fig. 3.38 were determined by the manual classification of 552 objects found within the cross-section of a single sample.[22] The segmented objects from future images can then be classified automatically by use of this chart. (Also, the proportion of objects found within each set gives an indication of the quality of sample preparation.) However, the minor-axis length decision lines are dependent on the fibre diameter distribution and must be adjusted to reflect the particular sample being analysed. On the other hand, the fit factor is scale independent, making these decision lines applicable to any sample.

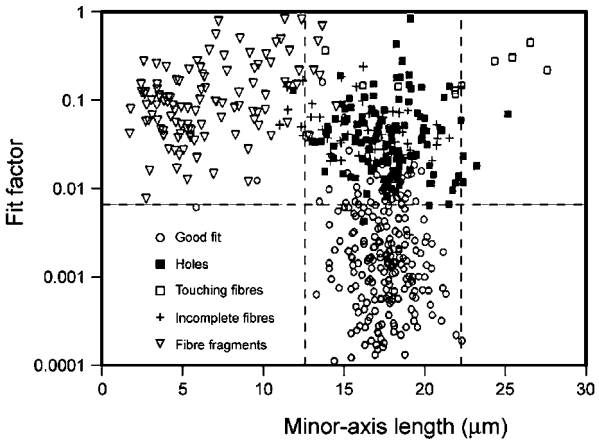


Figure 3.38 It is important to determine the quality of the fibre cross-section fitting. A fit factor may be defined and, depending on the values of the fit-factor against the fibre minor-axis length, one can identify (and then remove) fragments, incomplete fibres and unsuccessfully separated fibres in the dataset.

3.3.4 Stereology

In order to acquire unbiased data in any experiment, the sampling technique must be unbiased or, if a bias exists, it must be quantified and corrected. With the image analysis technique described here, fibres are only detected if they intersect the section plane and their intersection forms an ellipse. The probability that a fibre of length L will intersect a plane is dependent on its orientation with respect to the plane, as illustrated in Fig. 3.39. This introduces a bias, for example a fibre oriented at $\theta = 10^{o}$ is far more likely to be intersected by the plane than one of the same length oriented at $\theta = 50^{o}$.

Using simple geometry it can be shown that the relative probability of a fibre of length, L, and diameter, D, intersecting a plane oriented at an angle θ, to the fibre is given by,[45]

$$P(\theta) = L\cos\theta + D\sin\theta \qquad [3.26]$$

However, with the technique described in this case study, fibres are only counted if their cross-section forms an ellipse. Therefore the above formula must be adapted to give the probability of the intersection being a complete ellipse as follows:

$$P(\theta) = L\cos\theta - D\sin\theta \qquad [3.27]$$

Therefore, when calculating fibre orientation distribution functions, the data from every fibre must be divided by its intersection probability in order to remove this bias. However, calculation of the intersection probability for an individual fibre requires that the length of the fibre is known. Unfortunately, this information is not available to us, with one section of the fibre still embedded in the matrix and the other probably ground to a very fine dust! In practice, the mean fibre length $< L >$ is substituted for L in equation [3.27]. Furthermore,

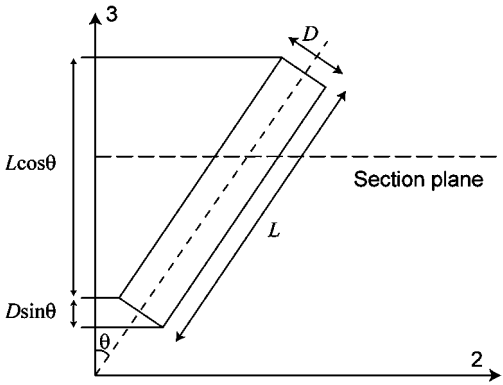


Figure 3.39 Orientation dependence on sampling – therefore a stereological bias is introduced.

with all but the shortest of fibres, the difference between a $\cos\theta$ and a $(\cos\theta - D \sin\theta/L)$ weighted fibre orientation distribution would be negligible.

3.3.5 Splitting of touching fibres

The fibre orientation measurement procedure described in Section 3.3.3 relies on the segmentation of the elliptical cross-sections within binary images. In high volume fraction samples these fibres are tightly packed and may appear to be touching within the digitised image. These touching fibres must be split before they are analysed using the moments technique. The perimeter of an ellipse is smooth and concave, however the perimeter of a pair of touching ellipses is sharply convex at the two points where they make contact. Therefore the perimeter curvature 'signature' can be used to identify touching fibres and their points of contact.

The edge pixels of an object are easily identified during region growing, i.e. any pixel that has neighbours which are below the intensity threshold is an edge pixel. Although, the number of edge pixels identified depends on whether 4-pixel or 8-pixel connectivity is being used, both will produce edge profiles that are suitable for perimeter tracing. In order to trace the perimeter of an object, an arbitrary starting point must be chosen. From this point the perimeter is traversed pixel-by-pixel until the starting point is reached once more. Figure 3.40 shows the edge pixels of two touching fibres and the perimeter tracing

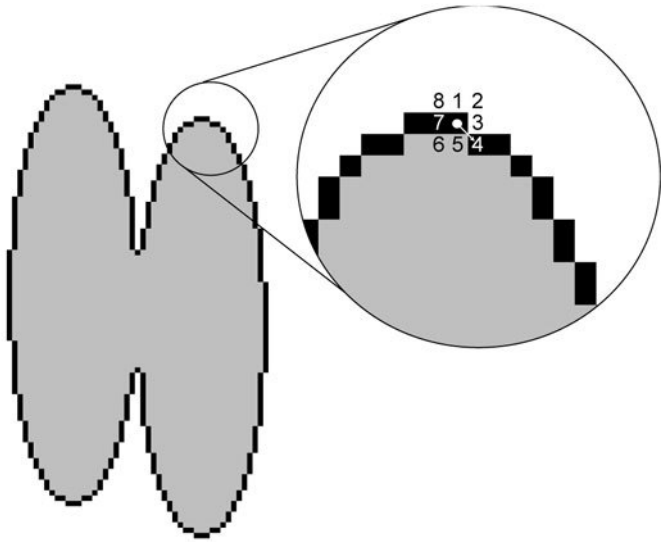


Figure 3.40 Two touching fibre cross-sections with their edges highlighted and the 8-pixel tracing route.

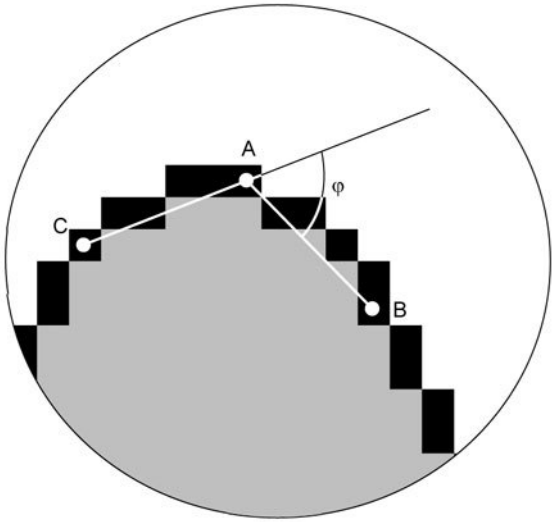


Figure 3.41 An illustration of how curvature may be measured around the fibre cross-section.

routine has been initiated at the indicated starting point. The neighbouring pixels are considered in the order illustrated until the first edge pixel is found. This will become the next 'active' pixel and the process is repeated from here. The orderly fashion in which neighbouring pixels are visited ensures that the perimeter is always traversed in the same direction and that edge pixels are not missed out. However, it is worth noting that this perimeter tracing procedure will fail if the object is less than three pixels wide. In this case, the edge pixels on each side of the object will adjoin. (A simple solution to this problem would be to dilate all the objects in the image, by adding a one pixel deep boundary around each, effectively thickening all of the objects.)

The local curvature at a point on an object's perimeter can be calculated by extending a secant to edge pixels a distance l away, as illustrated in Fig. 3.41. During perimeter tracing, points where the curvature becomes convex are recorded. When two fibres touch, there will be two convex perimeter points. When three fibres touch, there will be four convex points, and so on. These points are paired and lines plotted between them, splitting the fibres. This is illustrated in Fig. 3.42 where a large number of touching fibres have been found in a high packing fraction sample.

3.3.6 Fibre orientation distribution descriptors

The measurement technique described in this case study has the potential for producing large volumes of data. There are many ways in which 3D data can be described (see Section 3.2.4) and displayed.[46] One method that is in common

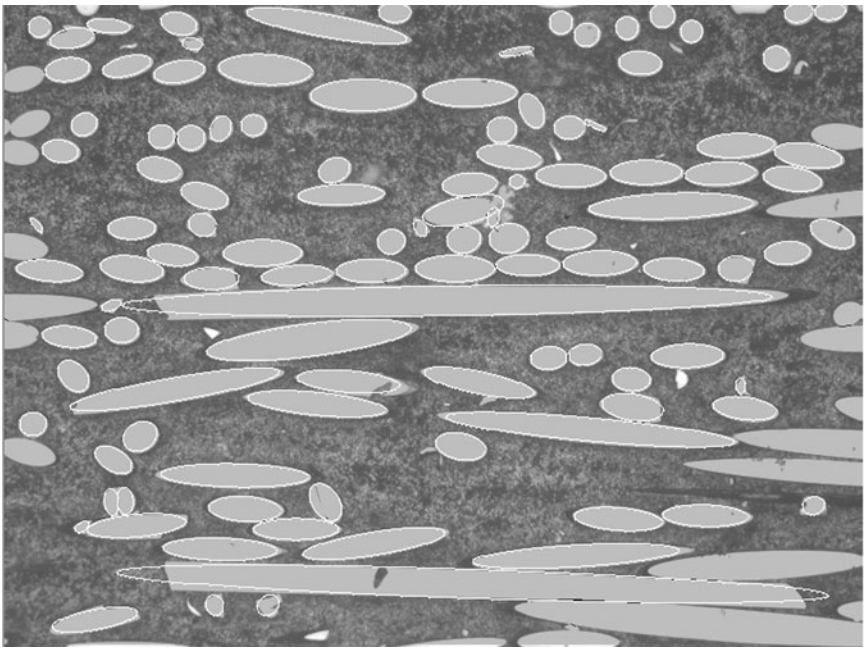


Figure 3.42 The frame (566 μm $\times$ 434 μm) shows the elliptical fitting to each fibre in the frame. Note how well the system copes with the effective separation of touching fibres. The fit factor will be poor for the two fibres which are lying in the section plane. The only fibres which do not have elliptical fits are those incomplete fibres at the edges of this frame, but they will be fitted after the XY stage has moved to another position.

use is 3D frequency plots as illustrated in Fig. 3.43. Clearly, if samples are being compared to each other, or if simulated orientation data are being evaluated, a more concise description of the data is required. There have been a number of different fibre orientation distribution descriptors suggested in the past and a review of some of the more popular ones can be found elsewhere.[47] Recently, orientation tensors have gained in popularity and are now the most widely used descriptor for fibre orientation.[31]

The spatial orientation of a single fibre is uniquely defined by two angles (θ_i, ϕ_i), alternatively it may be defined by a unit vector $\mathbf{p}$ which lies parallel to the fibre as follows,

$$\mathbf{p} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} \qquad [3.28]$$

A group of n fibres can be described by the orientation tensor, a, which is calculated by forming dyadic products of each individual fibre. The second-order orientation tensor is calculated as follows:

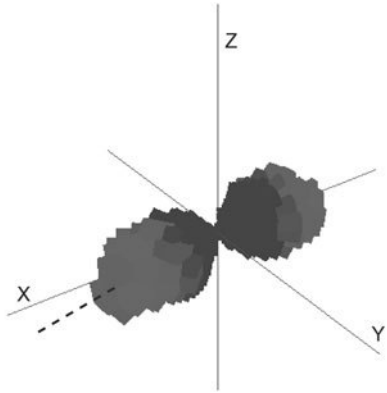


Figure 3.43 A graphical representation of the 3D probability distribution function of fibre orientations is shown. The symmetry of the probability function about the origin of this XYZ co-ordinate system is due to the ambiguity of the in-plane angle, ϕ. For this particular sample, there is a tendency for the fibres to be oriented within ten degrees of the X axis.

$$a_{ij} = \frac{1}{n}\left(\sum_{k=1}^{n} p_i^k p_j^k\right) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad [3.29]$$

A few common orientation distributions and their second-order orientation tensors are illustrated in Fig. 3.44.

The tensor description of fibre orientation has proved successful for a number of reasons. Firstly, unlike some of the earlier descriptors, the orientation tensor can describe distributions in three dimensions. In fact, as in Fig. 3.44, the dimensions can be chosen to suit the data. Secondly, the detail of the description can be improved by increasing the order of the tensor. Thirdly, composite fabrication simulation software often use tensors to describe other global material properties. Finally, the orientation tensor is easy to interpret.

3.3.7 Analysis of push-pull processing

To re-state the introduction to this case study, the most important feature of fibre-reinforced composites is the anisotropy they exhibit in their physical properties. Recently there has been a lot of interest in injection moulding as a means to fabricate composites. However, injection-moulded part performance, in terms of this anisotropy, is not as marked due to variations in fibre orientation throughout the thickness of the materials. Novel processing methods have been developed for optimising the moulding process to increase the fibre alignment. One such method is push-pull processing where the molten material is drawn

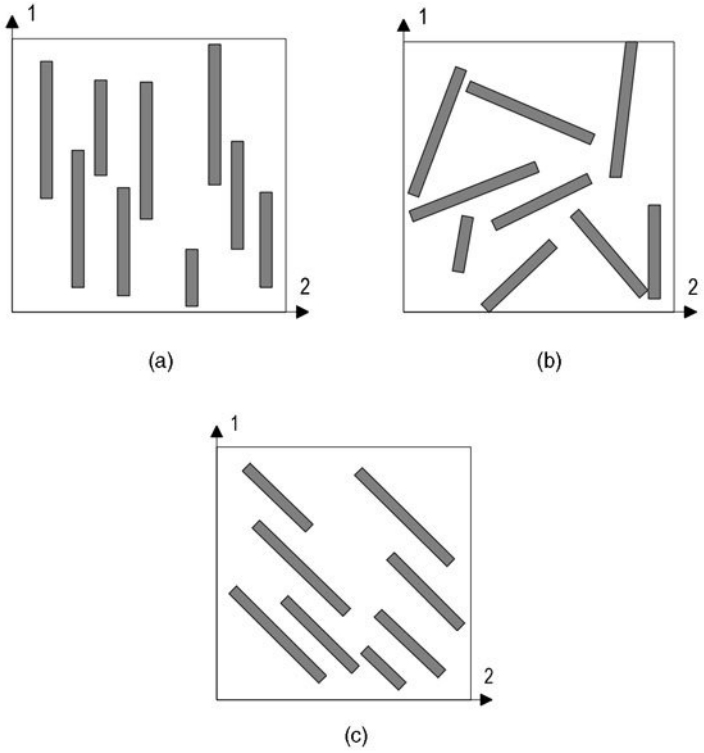


Figure 3.44 Three examples of 2D orientation distributions and their equivalent tensors: (a) unidirectional, $\mathbf{a} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, (b) isotropic, $\mathbf{a} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$ and (c) unidirectional, $\mathbf{a} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$.

back-and-forth through the mould.[48] Each layer of material solidifies under shear, resulting in a high degree of fibre alignment.

In order to evaluate the push-pull processing technique, a number of samples were manufactured, one in the conventional manner, i.e. a single injection of molten material, and the rest were push-pull processed with a varying number of strokes. Each sample was sectioned at the same location and the fibre orientation distributions at these locations were measured. Approximately 100,000 fibres were identified and analysed within each sample cross-section with an acquisition time of approximately four hours per sample. Figure 3.45 shows the variation in the a_{zz} component of the orientation tensor within two of the samples studied. The fibre orientation distributions were found to be essentially planar, $a_{yy} \approx 0$ and therefore the remaining diagonal component can be calculated as follows, $a_{xx} = 1 - a_{zz}$. The fibre orientation distribution in the conventional sample, Fig. 3.45(a), shows the skin-core-skin, which typifies

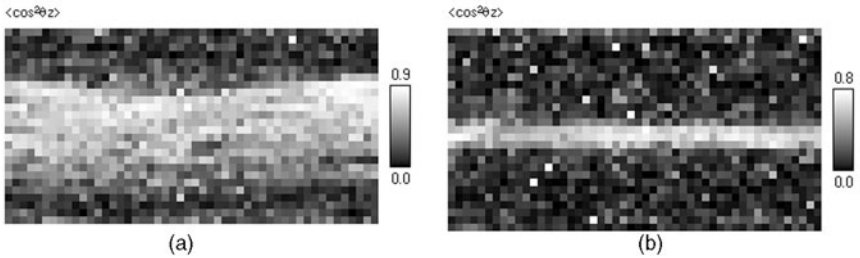


Figure 3.45 Large area scans (9.12 mm × 5.25 mm) through the thickness of an injection-moulded, fibre-reinforced composite are shown. The greyscale of each small square represents the average, $<\cos 2\theta_z>$ of the fibres within a 200μm × 200μm region. (Note that the few white squares, which are not connected to other white squares, are regions where no fibre was seen because of the low fibre packing fraction.) The conventional injection-moulded sample (a) shows a very thick central 'core' region where the fibres are oriented at right angles to those fibres in the outer 'skin' regions. However, a sample produced by the 8-stroke, 'push-pull' injection-moulding process (b) shows a very thin core region surrounded by thick skin regions.

injection-moulded composites. The other distribution in Fig. 3.45(b) demonstrates that the push-pull method does increase the degree of fibre alignment to the point that, in the 18-stroke sample, the skin-core-skin structure is barely visible.

3.3.8 Closing remarks

In this case study we have seen how automation and image processing can be used to collect large amounts of data from a sample, far more than is feasible with manual methods. The effect of physically sectioning the sample has been considered and the bias quantified. However, it is important to note that the method was adapted slightly from its original source[45] because of slight differences in measurement technique. Bias is always related to the measurement technique employed.

Although care has been taken to ensure that a large number of fibres are counted within each section plane, there are still a number of aspects of the measurement technique which have not been considered. Firstly, a fibre that is nearly parallel to the section plane will have a rectangular cross-section if both of its ends are intersected. The technique described will only measure the orientations of fibres which give an elliptical cross-section and hence there exists a cut-off angle after which no fibres will be detected, $\theta > \theta_c$ (i.e. an ellipse fitted to a rectangular cross-section will have a poor fit factor, $f > 0.1$). Therefore care must be taken when cutting a sample in order to minimise the number of fibres that fall within this range. Alternatively, the image processing routine can be modified to detect rectangular cross-sections.[45] Secondly, the accuracy of the fibre orientation measurement itself has not been considered.

The equation from which the fibre orientations are derived is non-linear and any measurement errors will be amplified when $a \approx b$ as is the case for fibres that are nearly perpendicular to the section plane. Eberhardt *et al.* have discussed the consequences recently in some detail.[12]

3.4 Case study: the measurement of fibre lengths in image fields

3.4.1 Summary

This case study describes the characterisation of a group of objects by a single parameter, the object's length and the application of large area scanning in order to analyse objects which are much larger than the field of view. A sample containing thousands of long fibres was analysed in this study. The fibres were often much longer than the field of view and were found to cross frequently. Both of these problems were overcome to give an accurate fibre length distribution.

3.4.2 Introduction

Both theory and experiment have proven that fibre-reinforced composites exhibit optimum performance when the fibres are long, i.e. extending throughout the length of the entire component. However, processing such components is both time-consuming and costly, often involving the manual lay-up of pre-prepared strips of fibres. The preparation of composites containing short fibres allows for more versatile processing routes such as injection moulding, where the molten polymer and fibre mix is forced into the mould. Although the length of the fibres in the raw material can be controlled, the injection moulding process is quite violent involving fibre breakage (which results in unpredictable fibre lengths in the finished component).

As the composite component is deformed, stresses are transferred from the matrix to the fibres embedded within it. The effectiveness of a fibre reinforcement depends on the extent of the stress transfer. Figure 3.46 shows the stress profile for a long fibre and a short fibre, both of which are surrounded by a matrix under the same tension (calculated using the shear lag model).[49] The stress profile of the long fibre has a plateau region, which indicates that the stresses are fully formed. However, the stresses within the short fibre are much lower and hence there is a greater load on the matrix which reduces the physical properties of the composite. Clearly, the distribution of fibre lengths within a short fibre composite will have a significant effect on the tensile properties of the component.

The physical properties of a composite component can be used to make some deduction of its fibre lengths. However, such indirect methods are unreliable, because other aspects of the composite microstructure may affect the same

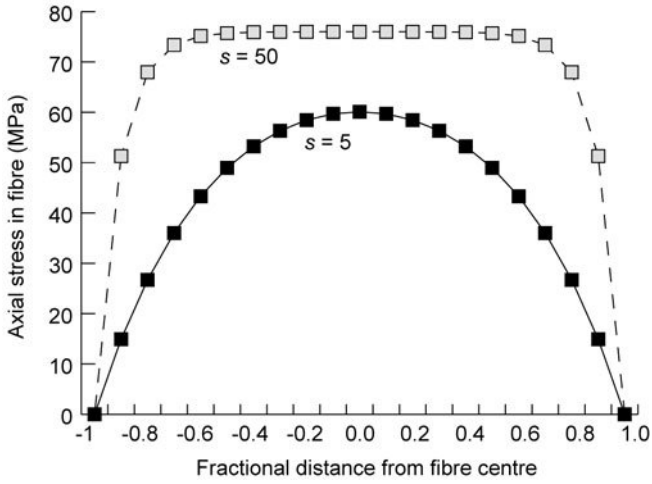


Figure 3.46 Stress profile for a long ($s = 50$) and a short ($s = 5$) fibre where s is the ratio of fibre length to fibre diameter. The calculations were performed for 30% volume fraction E-glass in nylon 66 resin and used the shear lag model of Cox.[49]

physical properties. The alternative approach is to use direct methods where the lengths of individual fibres are measured or inferred. Similar techniques, which are employed for the measurement of fibre orientation, can be extended to trace fibres over serial sections. These methods are time-consuming and prone to error. Stereology enables the inference of average fibre length from single sections[45] or pairs of sections[50] and are a practical alternative to the exhaustive serial sectioning method. However, if the fibre length distribution is required, these methods are inappropriate.

Fibres can be separated from the matrix by pyrolysing or dissolving the matrix and then their individual lengths can be measured. Figure 3.47 shows a transmission mode image of the fibres in a petri dish obtained from a short glass fibre-reinforced composite after pyrolysis (450 to 500ºC in the furnace for 3 hours to burn off the matrix). Note that if the lengths of carbon fibres are required, the fibres must be separated from the matrix by acid digestion. The lengths of individual fibres can be measured directly from these images after unwanted background artefacts have been removed by a suitable segmentation routine.

3.4.3 Line detection

With this particular problem, the segmentation of the image is trivial. Regions within the image are either fibre or background and these are clearly discernible when imaging the sample in transmission mode, as illustrated in Fig. 3.47. A suitable intensity threshold can be chosen and, as long as the microscope illumination remains unchanged, this threshold can be applied to all of the

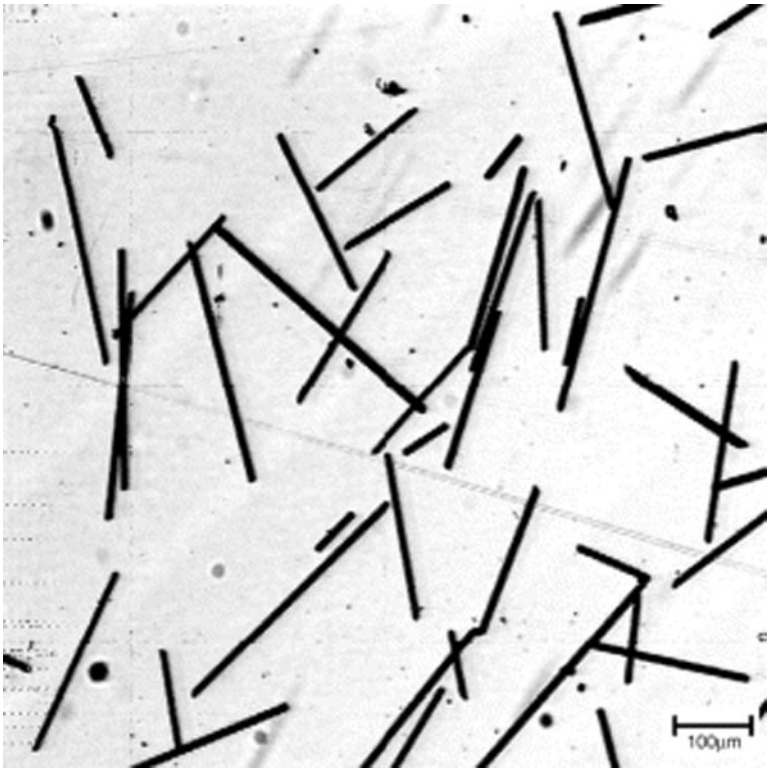


Figure 3.47 Transmission mode image of fibres that remain after the surrounding matrix has been pyrolised. Note the presence of both crossing fibres and connected fibres within the field of view.

images acquired for a particular sample. After segmentation, the individual objects within an image can be identified by region growing. If it were possible to prepare the sample to ensure that none of the fibres in the petri dish were crossing, the image processing task would be much simpler. During pixel connectivity, the extremes of each individual fibre could be recorded and their lengths determined directly. However, short of manually separating each fibre, this is not a feasible approach. The detected objects within the image field may be single fibres or a number of crossing fibres and hence further processing is required to separate them.

The identification of edges of objects and the discrimination of crossing features is a problem which is frequently addressed within the field of machine vision. For example, if a robot images an unknown object, it could recognise it as a cube by detecting its edges and comparing them against a known cube geometry.[51] The task of automatically identifying and separating fibres can be approached in a similar way, with each fibre described by the straight lines that form its boundaries. If there are a number of crossing fibres, the object boundary

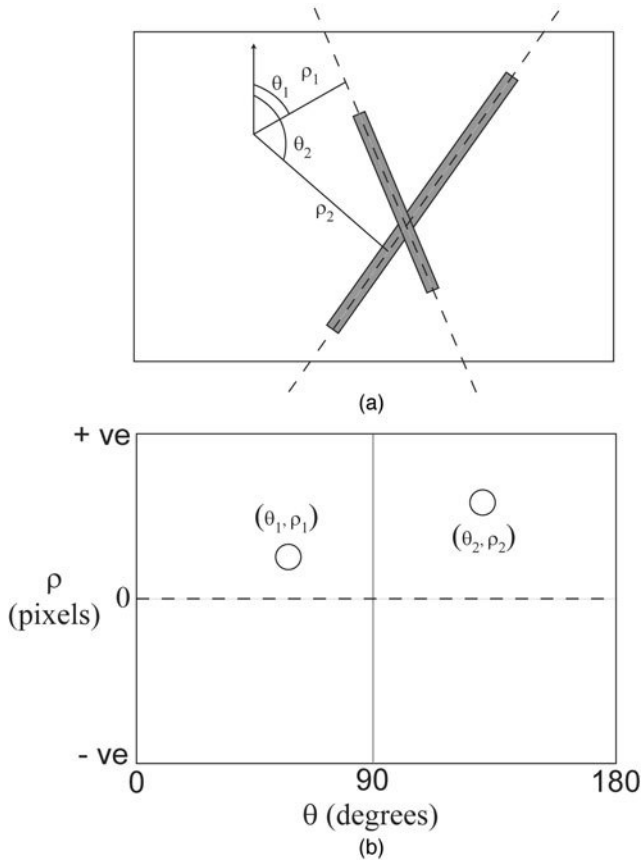


Figure 3.48 (a) The coordinate system used to characterise the fibre's position and orientation and (b) clusters in (θ, ρ) space.

will be made up of a number of straight-line segments. However, it should be possible to determine to which fibre a particular segment belongs (considering its position and orientation).

A line of infinite length can be described as follows,

$$x\cos(\theta) + y\sin(\theta) = \rho \qquad [3.30]$$

where ρ is the line's distance from the origin and the angle θ is the orientation of the line. Figure 3.48(a) illustrates an image containing two crossing fibres, which the region growing technique has identified as a single object. Ignoring the short-line segments which define the ends of each fibre, this object has a perimeter made from a total of eight straight-line segments. If each of these segments is described using the parametric description given in equation [3.30], the four line segments belonging to each fibre have the same orientation θ, and have a distance from the origin of ρ or $\Delta\rho + \rho$, where $\Delta\rho$ is the thickness of the

fibre. This technique of identifying clusters in θ and ρ (see Fig. 3.48(b)), can be used to separate large numbers of crossing fibres and is the basis for the fibre length measurement technique described below.

3.4.4 Method

The method described here uses binary image processing in order to locate individual fibres or groups of fibres. The edge pixels of each object are identified and the perimeter traced. The perimeter of each object is split into straight-line segments, which are described in terms of their distance from the origin and their orientation. These properties are used to identify which line segments belong to the same fibre. Finally, the length of each fibre is determined from the extent of these grouped line segments.

The transmission mode image shown in Fig. 3.47 shows excellent contrast between the fibres and the image background, making the selection of an appropriate threshold intensity very straightforward. All the objects within the image were located using the region growing technique described in Section 2.8.6. In the previous case study, the cross-sections of touching fibres were separated by locating points on the object perimeter which were convex (see Section 3.3.5). In this study, multiple crossing fibres are commonplace and a similar method is employed to split the straight-line segments that make up its perimeter. Figure 3.49 shows a pair of fibres that are crossing. The object perimeter has been traced and the curvature at each perimeter pixel calculated. The object's curvature 'signature' illustrated in Fig. 3.49 enables the identification of straight-line segments on the object's perimeter.

As the straight-line segments are identified, θ and ρ are stored as given by equation [3.30]. However, this equation defines an infinitely long line, whereas we are interested specifically in the length. Therefore the start and end points of each line segment are also stored. Clusters in θ and ρ are identified in order to group line segments from individual fibres. The values are often found in two clusters, (θ, ρ) and $(\theta, \rho + \Delta\rho)$ where $\Delta\rho$ is the fibre width, the mean value of θ and ρ is determined, giving the fibre centre line. By examining the start and end points and these clustered line segments, the two most distant points can be found which give the length of the respective fibre.

So far, the analysis has assumed that all the fibres within the image are perfectly straight, but with glass fibres this is not always the case. Slight curvature is often observed when the total fibre length is in the region of a few millimetres. This residual fibre curvature is unlikely to result in significant peaks in the object curvature profile illustrated in Fig. 3.50 and therefore its curved sides will be approximated by straight lines. However, this will have an effect on the fibre length measurement, with the lengths of curved fibres consistently under-estimated. A solution to this problem is to limit the maximum length of any one line segment, as illustrated in Fig. 3.51. The θ and ρ values for the line segments

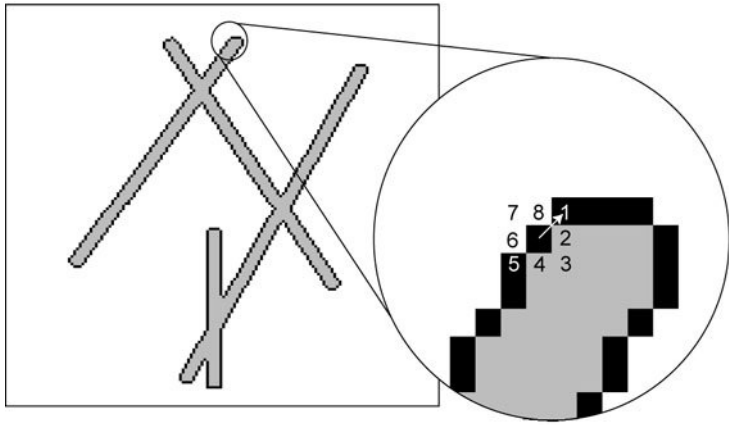


Figure 3.49 A schematic diagram showing crossing fibres and the derivation of the perimeter angle.

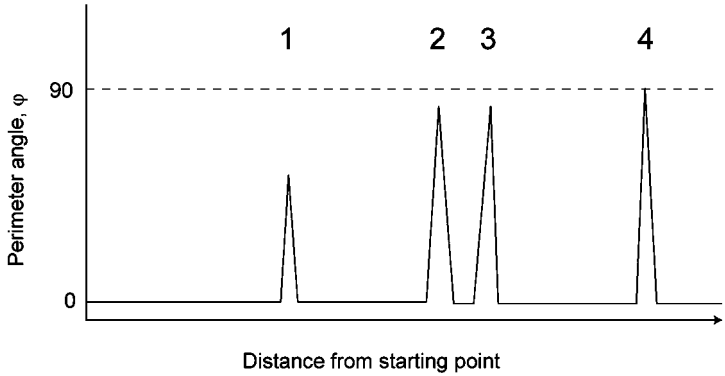


Figure 3.50 Typical curvature signatures observed for crossing fibres.

of this curved fibre will not form such a tight cluster. However, they are usually separable from those of crossing fibres. In the case of curved fibres, the total fibre length is calculated by summing the lengths of the adjoining line segments.

A suitable magnification must be chosen that ensures each fibre is at least a few pixels wide (to avoid the need for dilation as described earlier). With the Leeds system, a field of view of 1.2 mm × 1.1 mm is obtained with a magnification of 16.5×. At this magnification the width of a typical fibre, approximately 10 μm, will occupy 5 pixels (which provides sufficient resolution). However, the maximum length of fibre detectable within the image frame is approximately 1.6 mm. Fibre length measurements taken from a number of single image frames would therefore exhibit a considerable length bias. If one could be sure that the fibre lengths were in the region of 300 μm, the 'disector' principle discussed in Chapter 2 could be applied to the problem. However, the

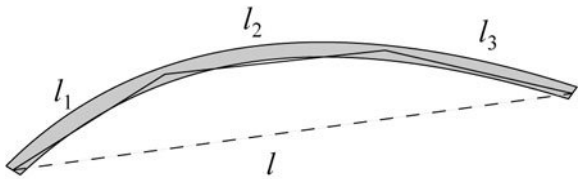


Figure 3.51 The length of a curved fibre is determined by summing the lengths of a number of short straight-line segments.

crossing fibres may occupy much larger regions, making it almost impossible to predict the exact bias.

3.4.5 Results

During the development of any automated image processing routine, it is essential that the results be tested against some standard. This may involve analysing a sample where the parameters being studied are known *a priori*, either through special sample preparation or comparison to another measurement technique. Alternatively, the results can be compared to those obtained by repetitive manual measurements.

A composite sample containing unidirectional, continuous glass fibres was cut to a length of approximately 5 mm, pyrolised and the fibres dispersed on a petri dish. A region measuring 70 mm × 70 mm was scanned and the lengths of approximately 2000 fibres were determined. The resulting length distribution illustrated in Fig. 3.52 shows a peak at approximately 5.2 mm with very little deviation beyond 0.5 mm, clearly demonstrating that the image processing algorithms are performing correctly.

Finally, to demonstrate the technique's performance on a sample of unknown fibre length distribution, an injection-moulded, short fibre composite sample was analysed. A section was cut through the sample and the skin and core regions identified and separated. After pyrolysis the two regions were analysed separately with the results shown in Fig. 3.53. The plot shows that the fibres are slightly longer in the core region than in the skin. This can be explained by the higher shear stresses that are present in the skin regions during mould filling leading to additional fibre breakage.

3.5 Latest developments

3.5.1 Pattern matching between successive 2D sections

It is a small logical step from analysing a single, polished 2D section to attempt to take a number of images on a set of polished sections and, by pattern matching between sections, to reconstruct fibres or voids in 3D. However the

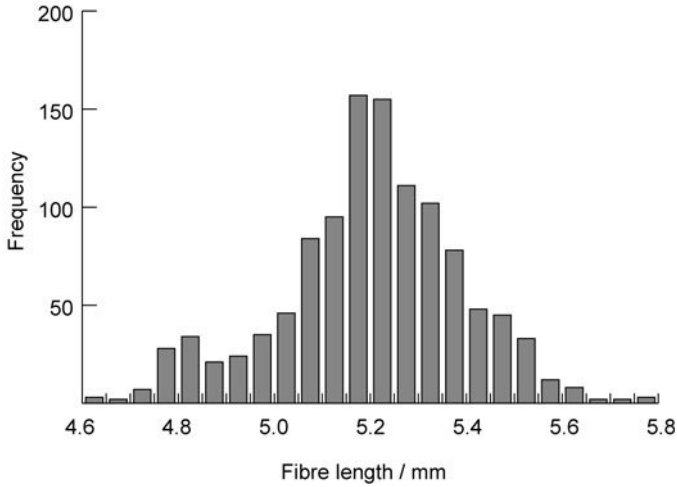


Figure 3.52 The frequency distribution of fibre lengths for a 5 mm section of a unidirectional material. The fibre lengths are distributed approximately normally about a mean length of 5.2 mm, as might be expected.

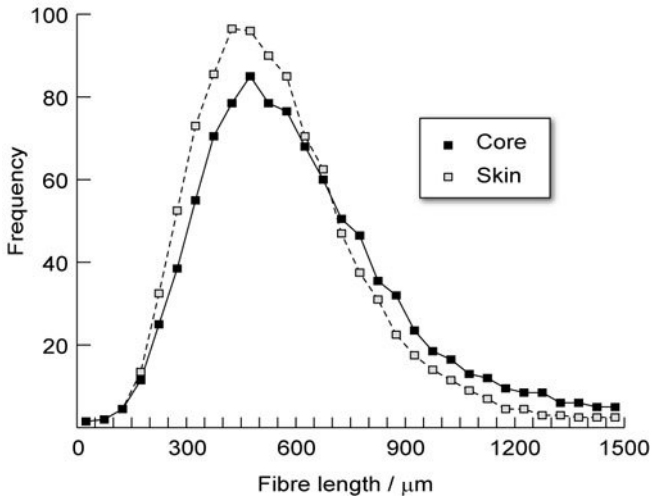


Figure 3.53 The fibre length frequency distribution within the skin and core regions of an injection-moulded composite, showing a large spread in fibre lengths. When normalised, it is clear that there are more of the longer fibres in the core region than the skin region.

practicalities are non-trivial, which probably explains why literature searches have only uncovered a few published papers on this topic!

There are many published papers on pattern recognition and pattern matching between two or more complex images (see, for example, Merickel,[53] Hibbard *et al.*[54]). However, the authors are only aware of a few attempts to successively

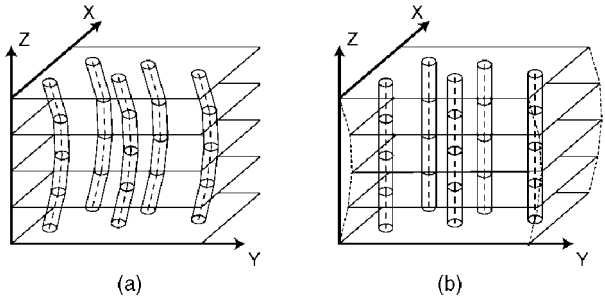


Figure 3.54 An illustration of Paluch's scheme for deducing fibre curvature from a series of physical sections. (a) If the curved fibres are synchronised spatially, as shown, (b) Paluch's procedure tends to straighten out the fibres.

section a composite sample and then to reconstruct the sample in 3D by pattern matching the fibre images between consecutive sections. The only paper that has appeared in an international composites conference was by Paluch[55] and the Paluch technique has been discussed in Clarke *et al.*[56] In essence, Paluch attempted to deduce the waviness of unidirectional carbon fibres by processing a single XY image frame at high magnification and then removing another 20 μm of sample material by sectioning and polishing, before scanning the same region again. Paluch repeated the process 40 times, removing 20 μm between each scan. The problem was to achieve good registration between these single frames. His method was to minimise the 'random' fibre centre movements between planes, see Fig. 3.54(a). However, as shown in Fig. 3.54(b), if there are any signs of correlated waviness among the fibre dataset, his centroid technique tends to straighten out the fibres, i.e. under-estimate the real waviness of the fibres! There is also the problem of knowing, to sub-micron precision, how much material, Δz, has been removed between sections.

Unlike Paluch, who tried to match a whole set of isolated, high spatial resolution, small area, X–Y frames, the high resolution, large area scanning capability of the Leeds 2D image analyser can be used to improve the pattern matching process significantly. The practical problem was to have good quality fibre images and excellent registration between fibre centres. Our aim was to remove the ambiguity in the in-plane angle, ϕ, by successful pattern matching between two sections and then to investigate the technique further for the systematic sectioning and 3D reconstruction of opaque fibre composites.

Consider an X–Y area of 5 mm $\times$ 5 mm, which was scanned (section A) and the fibre orientation information derived. After removing some material from the sample and polishing, the same region was scanned again (section B). Due to removing and replacing the sample on the microscope stage, there will be a small translational movement (Δx, Δy) and rotational difference (Ψ) between the coordinates within the two scans, as shown in Fig. 3.55.

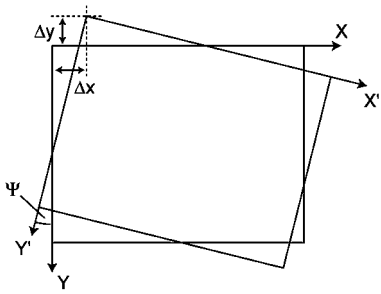


Figure 3.55 If physical sections are taken, one section can be considered different from the next by an offset, Δx and Δy, and an angular rotation, Ψ. Hence, when scanning over a large area, the fibre centre coordinates must be transformed between the X, Y and X′, Y′ coordinate systems.

The link between corresponding points in the section A (x, y) coordinate system and those in the section B (x', y') coordinate system is given by:

$$x'_i = (x_i \cos \Psi - y_i \sin \Psi) + \Delta x \quad [3.31]$$

$$y'_i = (x_i \sin \Psi + y_i \cos \Psi) + \Delta y \quad [3.32]$$

In pattern matching problems, it is important to identify 'control points', i.e. features on one section that are clearly present on the next section. If n control points that appear in both sections can be identified, the best fit set of parameters can be found by minimising the sum of squared errors. The partial derivatives with respect to Δx, Δy and Ψ are set to zero to derive a linear set of equations which can be solved.

$$\cos \Psi = \frac{n \sum_i^n (y_i y'_i + x_i x'_i) - \sum_i^n x_i \sum_i^n x'_i - \sum_i^n y_i \sum_i^n y'_i}{C} \quad [3.33]$$

$$\sin \Psi = \frac{n \sum_i^n (x_i y'_i - x'_i y_i) + \sum_i^n y_i \sum_i^n x'_i - \sum_i^n x_i \sum_i^n y'_i}{C} \quad [3.34]$$

$$\Delta x = \frac{-\sum_i^n x_i \sum_i^n (y_i y'_i + x_i x'_i) + \sum_i^n y_i \sum_i^n (x_i y'_i - x'_i y_i) + \sum_i^n (x_i^2 + y_i^2) \sum_i^n x'_i}{C} \quad [3.35]$$

$$\Delta y = \frac{-\sum_i^n y_i \sum_i^n (y_i y'_i + x_i x'_i) - \sum_i^n x_i \sum_i^n (x_i y'_i - x'_i y_i) + \sum_i^n (x_i^2 + y_i^2) \sum_i^n y'_i}{C} \quad [3.36]$$

where

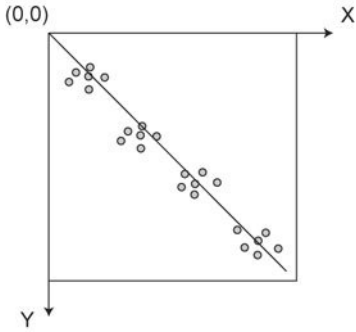


Figure 3.56 In order to deduce the offsets and rotation angle between the two image fields, one sensitive technique would be to consider the location of fibre clusters along the diagonal line of the scanned area.

$$C = n\sum_{i}^{n}(x_i^2 + y_i^2) - \left[\sum_{i}^{n} x_i\right]^2 - \left[\sum_{i}^{n} y_i\right]^2 \qquad [3.37]$$

In aligned fibre composites, sections would be taken normal to the preferred fibre direction and hence the only objects that can be used as control points are circular fibre images. As the 5 mm × 5 mm section plane may contain more than 15,000 fibre images, it was decided to define n clusters of six nearest neighbour fibres along the section diagonal (see Fig. 3.56). It was assumed that, over a fibre distance of 10 or 20 μm, the fibres within the cluster would retain their relative positions.

Each cluster selected in section A is characterised by the set of angles α_n and inter-fibre image distances l_n between members of the cluster, as shown in Fig. 3.57. These data produce an orientation independent definition of each cluster, which aids the unambiguous identification of cluster control points in section B. Hence, the translational and rotational differences between sections A and B may be deduced. Then, the centre coordinates of all fibre images on section B may be transformed to align with the coordinate system for section A. The first results obtained by this procedure have been described elsewhere.[40]

3.5.2 2D–3D optical microscopy

Clearly considerable effort (and patience) is required to undertake the sectioning of a sample hundreds (or thousands) of times in order to create a 3D reconstruction of the mesostructure. It would be much better if the 3D mesostructure of composite materials could be obtained from one single polished section. The Leeds approach to this problem will be obvious when the confocal laser scanning microscope is discussed in the next chapter. However, there is another optical microscopy approach which has been published recently – the optical sectioning method (OSM) – which can be used for transparent

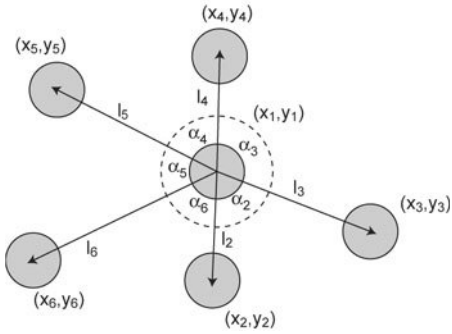


Figure 3.57 The key issue for registration of objects within images is to define some control points or easily recognisable features, which are common to both fields. The centre of a small cluster of fibres can often act as a suitable control point.

composites and which has yielded real-time image data during the processing of glass fibre-reinforced material.

McGrath and Wille's technique[38] uses 0.1 wt% opaque carbon 'tracer' fibres within a 30 wt% glass fibres in thermoplastic composites whose matrix refractive index is a close match to the refractive index of the reinforcing glass fibres. This OSM technique can focus on the tracer fibres, which are at different depths within this specially prepared composite sample. By controlling the Z movement of the objective lens, the focal plane moves within the sample. Unlike the Paluch technique discussed above, these focal planes are guaranteed parallel and perfectly registered, enabling unambiguous correlation of fibre locations in 3D.

Spatial resolutions of a few μm in δx and δy and approximately 10 μm in δz have been achieved. This matching refractive index requirement limits the range of matrix materials that can be investigated, but the technique can obtain 3D fibre orientation distributions from significant depths (1.3 mm has been quoted) and at high processing rates (between 1 and 10 $mm^3 min^{-1}$). An overview of their technique is shown in Fig. 3.58. A transparent window in the side of a mould has enabled them to explore the movement of fibres during processing and the measurement errors associated with their technique. Note that this technique does not have the same sub-micron spatial resolution as the CLSM technique discussed in the next chapter.

The latest technique for 3D reconstructing from 2D section planes has been advertised by the 3dRESOLVE website. At their website, they describe how they pot their samples in black epoxy resin (to obtain better contrast between the matrix and the material of the specimen). By successive 2D microtoming, imaging and pattern matching, they then build up the full 3D reconstruction using thousands of sectioned planes separated by only tens of microns. They call their procedure digital volumetric imaging.

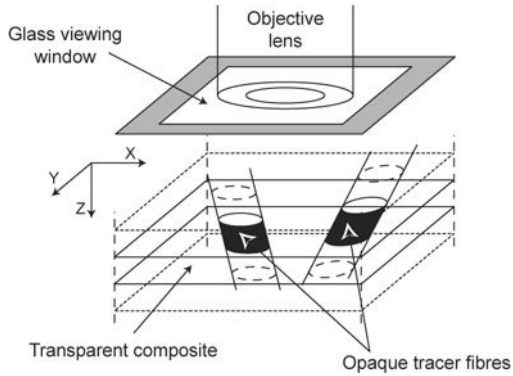


Figure 3.58 Schematic of McGrath and Wille's optical sectioning method, which allows them to optically section within a significant volume, using tracer fibres for better contrast and larger depths of view.

3.6 References

1 Piggott M.R. (1994), The effect of fibre waviness on the mechanical properties of unidirectional fibre composites, *Proc. Conf. on Mesostructures and Mesomechanics in Fibre Composites*, ed M R Piggott, Univ. of Toronto, pp 145–158.

2 Mlekusch B., Lehner E.A. and Geymayer W. (1999), Fibre orientation in short-fibre-reinforced thermoplastics I: Contrast enhancement for image analysis, *Comp. Sci. Tech.*, **59**, 543–545.

3 Haynes R. (1984), *Optical Microscopy of Materials*, Blackie & Son Ltd, pp 94–110.

4 Fynn G.W. and Powell J.A. (1983), *Cutting and Polishing Optical and Electronic Materials*, 2nd edn, Adam Hilger, Bristol.

5 Bay R.S. and Tucker III C.L. (1992), Stereological measurement and error estimates for three-dimensional fibre orientation, *Polym. Eng. Sci.*, **32**(4), 240–253.

6 Yurgartis S.W. (1987), Measurement of small angle fibre misalignments in continuous fibre composites, *Comp. Sci. Tech.*, **30,** 279–293.

7 Hine P.J., Duckett R.A., Davidson N.C. and Clarke A.R. (1993), Modelling of the elastic properties of fibre reinforced composites, 1: orientation measurement, *Comp. Sci. Tech.*, **47**, 65–73.

8 Toll S. and Andersson P-O. (1991), Microstructural characterisation of injection moulded composites using image analysis, *Composites*, **22**(4), 298–306.

9 Fischer G. and Eyerer P. (1988), Measuring the spatial orientation of short fibre reinforced thermoplastics by image analysis, *Polym. Comp.*, **9**(4), 297–304.

10 Clarke A.R., Davidson N.C. and Archenhold G. (1993), Measurements of fibre directions in reinforced polymer composites, *J. Microsc-Oxford*, **171**, 69–79.

11 Mlekusch B. (1999), Fibre orientation in short-fibre-reinforced thermoplastics II: Quantitative measurements by image analysis, *Comp. Sci. Tech.*, **59**, 547–560.

12 Eberhardt C.N., Clarke A.R., Vincent M., Giroud T. and Fluoret S. (2001), Fibre orientation measurements in short-glass-fibre composites II: a quantitative error estimate of the 2D image analysis technique, *Comp. Sci. Tech.*, **61**, 1961–1974.

13 Hine P.J., Duckett R.A., Davidson N. and Clarke A.R. (1993), Modelling of the elastic properties of fibre reinforced composites, 1: orientation measurement, *Comp. Sci. Tech.*, **47**, 65–73.

14 Gundersen H.J.G., Bagger P., Bendtsen T.F., Evans S.M., Korbo L., Marcussen N., Moller A., Nielsen K., Nyengaard J.R., Pakkenberg P., Sorensen F.B., Vesterby A. and West M.J. (1988), The new stereological tools: disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis, *APMIS 96*, 857–881.
15 Cruz-Orive L.M. and Weibel E.R. (1990), Recent stereological methods for cell biology: brief survey, *Amer. J. Physiol.*, **258**, L148–L156.
16 Howard V. (1990), Stereological techniques in biological electron microscopy, Chapter 13, *Biophysical Electron Microscopy*, Academic Press.
17 Möginger B. and Eyerer P. (1991), Determination of the weighting function $g(\beta, r, v)$ for the fibre orientation analysis of the short fibre reinforced composites, *Composites*, **22**, 394–398.
18 Mattfeldt T., Clarke A.R. and Archenhold G. (1994), Estimation of the directional distribution of spatial fibre processes using stereology and confocal scanning laser microscopy, *J. Microsc-Oxford*, **173**(2), 87–101.
19 Guild F.J. and Summerscales J. (1993), Microstructural image analysis applied to fibre composite materials: a review, *Composites*, **24**, 383–393.
20 Fakirov S. and Fakirova C. (1985), Direct determination of the orientation of short glass fibers in an injection-molded poly(ethylene terephthalate) system, *Polymer Composites*, **6**, 41–46.
21 Yurgartis S.W., Morey K. and Jortner J. (1993), Measurement of yarn shape and nesting in plain-weave composites, *Comp. Sci. Tech.*, **46**, 39–50.
22 Davidson N.C. (1993), Image analysis for fibre orientation measurement, PhD Thesis, Department of Physics and Astronomy, University of Leeds.
23 Clarke A.R., Davidson N.C. and Archenhold G. (1990), A multitransputer image analyser for 3D fibre orientation studies in composites. In H.Y. Elder (ed.), *Trans. Royal Microscopical Soc.*, Adam Hilger Publishers, **1**, 305–309.
24 Clarke A.R., Davidson N.C. and Archenhold G. (1991), A large area, high resolution, image analyser for polymer research, *Proc. Int. Conf. Transputing '91*, **1**, 31–47.
25 Stobie R.S. (1986), The COSMOS image analyser, *Pattern Recogn. Lett.*, **4**(5), 317–324.
26 Wu W-Y. and Wang M-J.J. (1993), Elliptical object detection by using its geometric properties, *Pattern Recognition*, **26**(10), 1499–1509.
27 O'Donnell B. and White J.R. (1991), Young's Modulus variation within glass-fibre-filled nylon 6,6 injection mouldings, *Plastics & Rubber Institute, Proc. 1st Conf. on Deformation and Fracture of Composites*, UMIST Manchester, 24.1–24.6.
28 Ward I.M. (1977), The measurement of molecular orientation in polymers by spectroscopic techniques, *J. Polym. Sci.: Polymer Symposium*, **58**, 1–21.
29 Ward I.M. (1962), Optical and mechanical anisotropy in crystalline polymers, *Proc. Phys. Soc.*, **80,** 1176–1188.
30 Hine P.J., Duckett R.A. and Ward I.M. (1993), Modelling the elastic properties of fibre reinforced composites: II theoretical predictions, *Comp. Sci. Tech.*, **49**, 13–21.
31 Advani S.G. and Tucker C.L. (1987), The use of tensors to describe and predict fiber orientation in short fiber composites, *Polym. Eng. Sci.*, **31**(8), 751–784.
32 Ludwig H-C., Fischer G. and Becker H. (1994), A quantitative comparison of morphology and fibre orientation in push-pull processed and conventional injection moulded parts, *Proc. Conf. on Mesostructures and Mesomechanics in Fibre Composites*, ed. M.R. Piggott, Univ. of Toronto, pp 223–232.

33 Taya M., Muramatsu K., Lloyd D.J. and Watanabe R. (1991), Determination of distribution patterns of fillers in composites by micromorphological parameters, *JSME International Journal*, **34**(2), 80–91.
34 Yurgartis S.W. (1994), Techniques for the quantification of composite mesostructure, *Proc. Conf. on Mesostructures and Mesomechanics in Fibre Composites*, ed. M.R. Piggott, Univ of Toronto, pp 32–55.
35 Clarke A.R., Davidson N.C. and Archenhold G. (1995), The measurement and modelling of fibre directions in composites, *IUTAM Symposium on Microstructure-Property Interactions in Composite Materials*, ed. R. Pyrz, Kluwer Academic Publishers, pp 77–88.
36 Pyrz R. (1994), Quantitative description of the microstructure of composites. Part 1: morphology of unidirectional composite systems, *Comp. Sci. Tech.*, **50**, 197–208.
37 Camponeschi E.T. (1991), Lamina waviness levels in thick composites and its effect on their compression strength, *Proc. Int. Conf. on Composite Materials, ICCM8*, Honolulu, pp 30-E-1–30-E-13.
38 McGrath J.J. and Wille J.M. (1995), Determination of 3D fiber orientation distribution in thermoplastic injection molding, *Comp. Sci. Tech.*, **53,** 133–143.
39 Denton D.L. and Munson-McGee S.H. (1985), The use of X-radiographic tracers to measure fiber orientations in short fiber composites, *High Modulus Fiber Composites in Ground Transportation and High Volume Applications*, Wilson D.W. (ed.), ASTM, Philadelphia.
40 Davidson N.C., Clarke A.R. and Archenhold G. (1997), Large-area, high-resolution image analysis of composite materials, *J. Microsc-Oxford*, **185**(2), 233–242.
41 Wetherhold R.C. and Scott P.D. (1990), Prediction of thermoelastic properties in short-fiber composites using image analysis techniques, *Comp. Sci. Technol.*, **37**(4), 393–410.
42 Daul C., Graebling P. and Hirsch E. (1998), From the Hough transform to a new approach for the detection and approximation of elliptical arcs, *Comput. Vis. Image. Und.*, **72**(3), 215–236.
43 Yurgatis S.W. and Purandare M.N. (1991), Describing fiber spatial distribution in unidirectional composite materials, *Journal of Computer-Assisted Microscopy*, **3**(3), 117–125.
44 O'Connell P.A. and Duckett R.A. (1991), Measurements of fibre orientation in short-fibre-reinforced thermoplastics, *Comp. Sci. Tech.*, **42**, 329–347.
45 Zhu Y.T., Blumenthal W.R. and Lowe T.C. (1997), Determination of non-symmetric 3-D fiber-orientation distribution and average fiber length in short-fiber composites, *Journal of Composite Materials*, **31**(13), 1287–1301.
46 Fisher N.I., Lewis T. and Embleton B.J.J. (1998), *Statistical Analysis of Spherical Data*, Cambridge University Press.
47 Papathanasiou T.D. and Guell D.C. (1997), *Flow-induced Alignment in Composite Materials*, Woodhead Publishing Limited.
48 Becker H., Fischer G. and Muller U. (1993), Push-pull injection moulding of industrial products, *Kunstoffe German Plastics*, **83**(3), 3–4.
49 Cox, H.L. (1952), The elasticity and strength of paper and other fibrous materials, *Brit. J. Appl. Phys.*, **3**, 72–79.
50 Zak G., Haberer M., Park C.B. and Benhabib B. (2000), Estimation of average fibre length in short-fibre composites by a two-section method, *Comp. Sci. Tech.*, **60**, 1763–1772.

51 Sonka M., Hlavak V. and Boyle R. (1993), *Image Processing, Analysis and Machine Vision*, Chapman & Hall.
52 Davidson N.C. and Clarke A.R. (1998), Extending the dynamic range of fibre length and fibre aspect ratios by automated image analysis, *J. Microsc-Oxford*, **196**(2), 266–272.
53 Merickel M. (1987), 3D reconstruction: the registration problem, *Computer Vision, Graphics and Image Processing*, **42**, 206–219.
54 Hibbard L.S., Grothe R.A., Arnicar-Sulze T.L., Dovey-Hartman B.J. and Pages R.B. (1993), Computed 3D reconstruction of median-eminence capillary modules: image alignment and correlation, *J. Microsc-Oxford*, **171**(1), 39–56.
55 Paluch B. (1993), Analysis of geometric imperfections in unidirectionally-reinforced composites, *Proc. of European Conf. on Composite Materials, ECCM6*, Bordeaux, pp 305–310.
56 Clarke A.R., Archenhold G. and Davidson N.C. (1995), A novel technique for determining the 3D spatial distribution of glass fibres in polymer composites, *Comp. Sci. Tech.*, **55**, 75–91.

4

3D confocal laser scanning microscopy

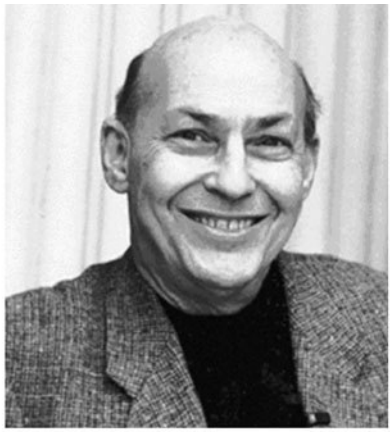
Any sufficiently advanced technology is indistinguishable from magic.
Arthur C. Clarke in *Profiles for the Future*

4.1 Principles of confocal laser scanning microscopy

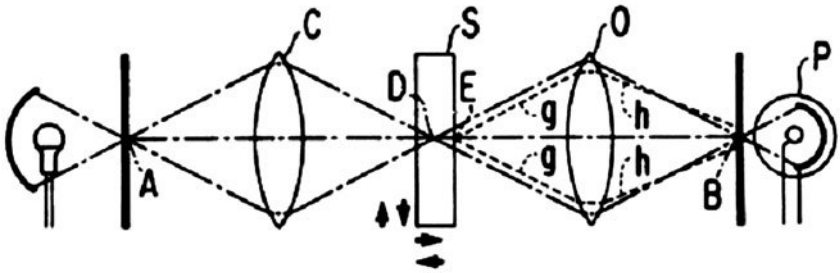
Confocal microscopy is a non-destructive sectioning technique, which enables the 3D reconstruction of material surfaces (surface topological studies) or the internal structure of semi-transparent specimens (sub-surface bulk studies). These spatial reconstructions by confocal microscopy entail either high spatial resolution ($\pm 0.2\mu$m) studies of regions within 5 μm to 200 μm of the specimen's surface (dependent on the opacity and heterogeneity of the sample) or, using air objective lenses, correspondingly lower resolution studies of regions within accessible depths of many millimetres. These upper limits to performance are determined essentially by the working depth of the high numerical aperture (NA) objective lens selected for the particular research study.

4.1.1 The first confocal microscope designs

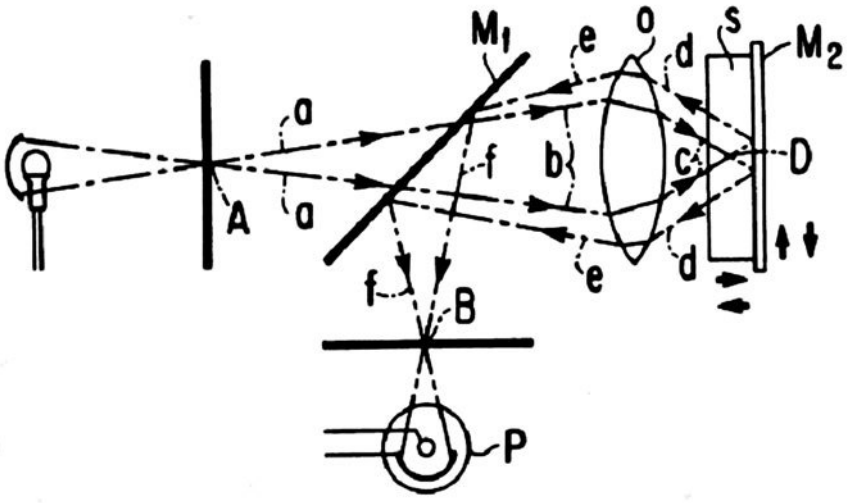
In 1957, Marvin Minsky, see Fig. 4.1(a), while working at Harvard University, filed his patent for 'a double-focussing stage scanning microscope'.[1] In so doing, he set the scene for the development of the confocal microscopes capable of 3D object reconstructions by optical sectioning, which are available today. Figure 4.1(b) shows Minksy's original 'double-focussing' design where light, emitted by a lamp, passes though a pinhole and is focussed onto the sample by a condenser lens. Light transmitted through the sample is focussed by an objective lens, through a second pinhole, onto a detector. The objective and condenser are confocally (having the same point of focus) aligned, hence the origin of the name 'confocal microscopy'. By simple ray tracing, it can be seen that light from any region, other than the focal point of both lenses within the sample, is rejected. A confocal microscope based on this design enables the formation of an image from a prescribed section of a thick translucent object, without the



(a)



(b)



(c)

Figure 4.1 (a) Portrait of Marvin Minsky and his original confocal designs in both (b) transmitted light and (c) epi-illumination modes of operation, from the original microscopy apparatus patent.[1]

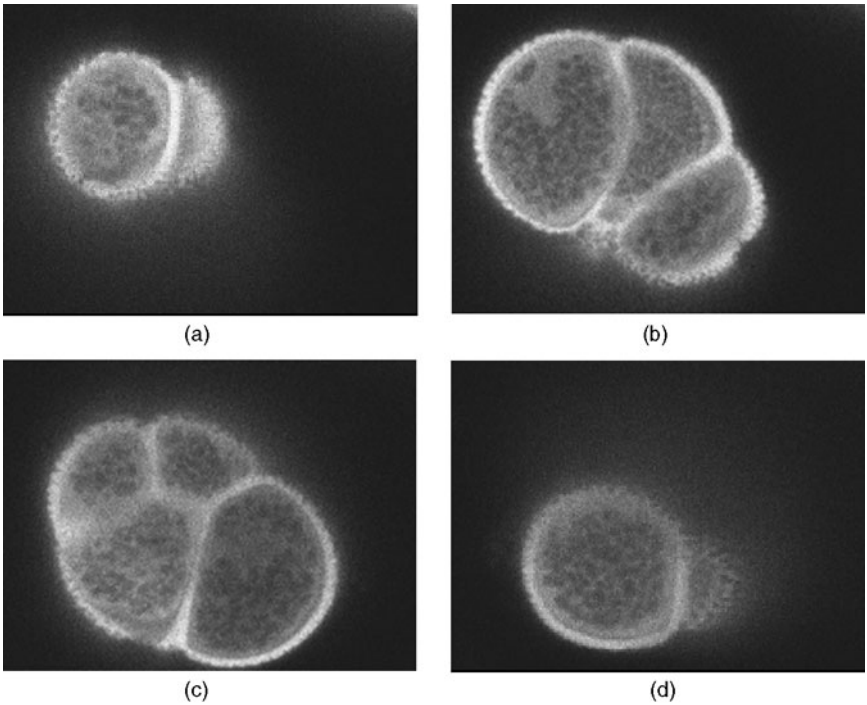


Figure 4.2 Four optical sections of a pollen grain each separated by a distance of $\Delta z = 10\,\mu$m.

presence of out-of-focus information from the surrounding planes. Hence, confocal images are frequently described as optical sections, enabling rapid 3D reconstruction without the need to physically section the sample. This property is illustrated in Fig. 4.2, where a pollen grain sample has been non-destructively, optically sectioned. Comparable reconstructions from a conventional microscope would require microtoming of the sample into a large number of thin slices, imaging each separately and finally registering them between sections (as discussed in Chapter 3).

With the use of higher NA objective lenses, the confocal alignment of both the condenser and objective lenses becomes more difficult (due to the diminishing focal spot size). An alternative approach is an epi-illumination confocal system, as illustrated in Fig. 4.1(c). Here, the same lens is used as both condenser and objective, ensuring confocal alignment (a lens is always confocally aligned with itself!) and this is now accepted as the classical form of a confocal microscope. This design also incorporates two pinholes; one in front of the illumination source and a second in front of the detector. The smaller the size of these apertures, the more stray light would be removed and the smaller the sample volume element that would be imaged.

Although Minsky referred to a bright, white light source in his original patent, the laser is now the universal light source employed within these confocal systems. The usual term to describe this type of microscope is either confocal laser scanning microscopy (CLSM) or, alternatively, confocal scanning laser microscopy (CSLM). In using a laser to illuminate the sample, the illumination aperture (labelled 'A' in Fig. 4.1(c)) can be removed, further simplifying the optics.

Both of the optical designs illustrated in Fig. 4.1 provide a confocal image of a single point within the sample. In order to form a 2D image of the focal plane, a means must be employed to move the illuminated point systematically over a region of the focal plane, effectively scanning the sample. In Minsky's original incarnation, the scanning was performed by physically moving the sample within the focal plane, on a mechanical XY stage. The main advantage of this fixed beam technique is that the beam does not deviate from the principal axis of the objective, effectively relaxing the requirements of this component and removing the need for highly corrected objective lenses. (Indeed Minsky noted that with his original design 'The usual microscope lens correction for coma, astigmatism, curvature of field, field distortion, and lateral chromatic aberration could be dispensed with'.) Furthermore, the magnification of the image is dictated entirely by the area of the region scanned, effectively de-coupling the microscope's magnification and field of view from those of the objective lens. In practice, this feature can be used to give a variable magnification range enabling the field of view to extend over many millimetres, which is far beyond the maximum field of view of standard objective lenses. The main drawback of the technique, which is common to all stage scanning microscope designs (some of which are discussed in Chapters 5 and 6), is the time required (typically tens of seconds) to scan a single image.

Most modern commercial confocal microscopes use a laser to create a focussed spot within the sample. An image is formed by scanning the laser beam through the objective lens in a raster pattern, as illustrated in Fig. 4.3. The reflected light is de-scanned using the same optics and focussed through a pinhole onto the detector. The speed of image collection is entirely dependent on the optical components which scan the laser beam. A large variety of scanning optics have been developed including Nipkow disk illumination, where scanning is performed by a rapidly spinning disk of small apertures,[2] oscillating-fibre illumination[3] and galvanometer-driven mirrors.[4] In all of these designs, the inertia of the moving parts is significantly lower than that of the stage assembly required in Minsky's design, resulting in a much faster scan rate. The use of galvanometer-driven mirrors appears to be the most popular in commercial microscope systems, offering an image acquisition rate of approximately 1–3 frames per second. However, the advantages of the stage-scanning design, namely relaxed lens requirements and de-coupled magnification, are unfortunately lost in all these scanning designs. Such is the desire for rapid imaging!

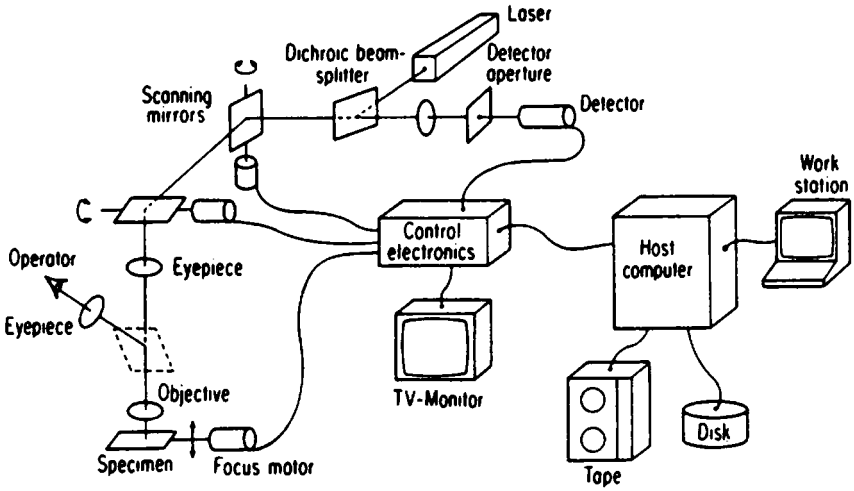


Figure 4.3 A schematic of a confocal laser scanning microscope (CLSM) (from Åslund *et al.*[4]).

Over the past fifteen years, many commercial confocal systems have appeared and there has been most interest (and hence more research papers have been published) in the biological and physiological, rather than the materials, research areas. If the reader is interested in a comprehensive treatment of biological applications of CLSM, there is a superb reference book edited by Pawley.[5] Also, a mathematically rigorous account of confocal optics and image formation can be found in the book by Wilson.[6] A number of groups have explored confocal microscopy for 3D surface topological studies in materials science, e.g. Gee.[7] Dutch groups have been active both in confocal design (e.g. Brakenhoff *et al.*[8]) and, also, in some of the earliest papers, which speculated on its potential use in composites research (e.g. Knoester and Brakenhoff[9] and Thomason and Knoester[10]).

Rather than 'reinventing the wheel', the emphasis of this chapter will be to illustrate typical pitfalls when operating CLSM systems in materials science. Those considerations that could improve the acquisition and the interpretation of images obtained by CLSM systems will be discussed. Before looking at specific case studies in materials science involving the CLSM, the foundations of confocal microscopy will be covered briefly in Section 4.1.2. The operational attributes of two commercial systems will be considered in Section 4.2: the Biorad MRC 500 and the Noran 'Odyssey' (because these have been the two CLSM systems that have been used by the authors over the past six years). Sections 4.3–4.5 will then deal with the practical issues of acquiring, interpreting and hence reconstructing the 3D structure of a sample.

4.1.2 The optics of confocal microscopy

A commonly used method for determining the resolution of a microscope involves the imaging of sub-resolution, point objects. The image of an infinitely small luminous object as formed by an objective lens is not itself infinitely small (due to the diffraction effects of the lens). Therefore, an Airy pattern, which consists of a central bright disk and progressively weaker rings, is observed. It can be shown that, when two equally bright points are separated by a small distance *d*, they may be resolved if the peaks of their Airy patterns are separated by the following distance:

$$r_{\text{Airy}} = \text{XY}_{res}^{conv} = 0.6\frac{\lambda}{\text{NA}_{\text{obj}}} \quad [4.1]$$

where r_{Airy} is the radius of the first dark ring in the Airy pattern, which may also be considered the lateral resolution, XY_{res}^{conv}, of the conventional microscope. Because the object is luminous, it is not necessary to include the NA of the condenser lens. This condition is known as the Rayleigh criterion and has been discussed previously in Chapter 1.

The confocal method of image formation results in a sharper peak in the central disk of the Airy pattern.[6] The intensity of the Airy pattern formed by a conventional microscope decreases by a factor of n^{-2}. However, with a confocal microscope, the intensity decreases by a factor of n^{-4}, where n is a measurement of lateral displacement in optical units. This clearly has an effect on the resolution of the microscope, giving an improvement in lateral resolution for the CLSM over the conventional microscope of ~1.5 when the Rayleigh criterion is applied. The Rayleigh criterion for a confocal microscope is therefore:

$$\text{XY}_{res}^{conf} = 0.4\frac{\lambda}{\text{NA}_{\text{obj}}} \quad [4.2]$$

which determines the lateral resolution, XY_{res}^{conf}, of the confocal microscope.

This small improvement in lateral resolution rarely offers any practical advantages to the microscopist but, to see the real advantages offered by confocal microscopes, the Airy pattern must be studied further and the axial resolution considered. The full 3D distribution of energy within the Airy pattern was calculated in 1956 by Linfoot and Wolf.[11] The 3D Airy pattern, termed the point spread function (PSF) of the microscope, dictates the observed appearance of an object. (The image formed is actually the convolution of the PSF and the object, see Section 2.8.4.) The PSF can be determined from the observation of sub-resolution fluorescent beads. The PSFs for a conventional and a confocal microscope are radially symmetrical and cross-sections of the measured intensity distribution for both microscopes are shown in Fig. 4.4. With a conventional microscope, the PSF is conical, extending both above and below the point object, whereas with a confocal microscope, it is cylindrical and does not extend as far in the axial direction.

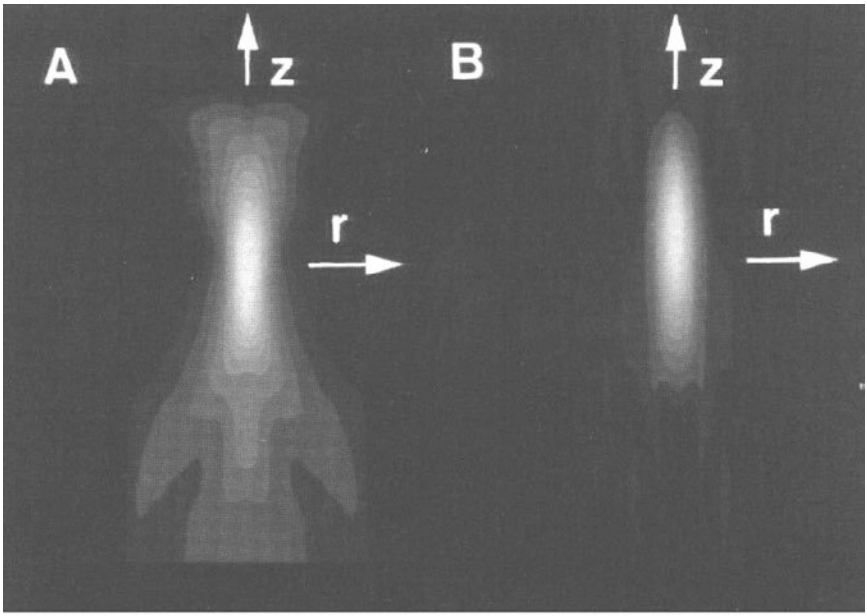


Figure 4.4 The experimentally determined point spread functions (PSFs) for (a) a conventional microscope and (b) a confocal microscope[5].

The Rayleigh criterion can also be applied to calculate the axial resolution of a microscope. For a conventional microscope, two point objects, separated by a small distance along the optical axis, can be resolved if the peaks of their Airy patterns are separated by the following distance:

$$Z_{res}^{conv} = 2\frac{\lambda}{\mathrm{NA}_{\mathrm{obj}}^2} \qquad [4.3]$$

where Z_{res}^{conv} is the axial resolution of the conventional microscope. In a confocal microscope, the change in energy distribution of the Airy pattern leads to an improvement in the lateral resolution of $\sim$1.4, with the Rayleigh criterion fulfilled for two points separated by the following distance:

$$Z_{res}^{conf} = 1.4\frac{\lambda}{\mathrm{NA}_{\mathrm{obj}}^2} \qquad [4.4]$$

where Z_{res}^{conf} is the axial resolution of the confocal microscope.

Although the confocal microscope offers a small increase in both lateral and axial resolutions, these do not account for the dramatic difference between confocal and conventional images. The appearance of out-of-focus material is determined by the shape of the PSF (at small distances from the point of focus). With a conventional microscope, the conical shape of the PSF, as shown in Fig.

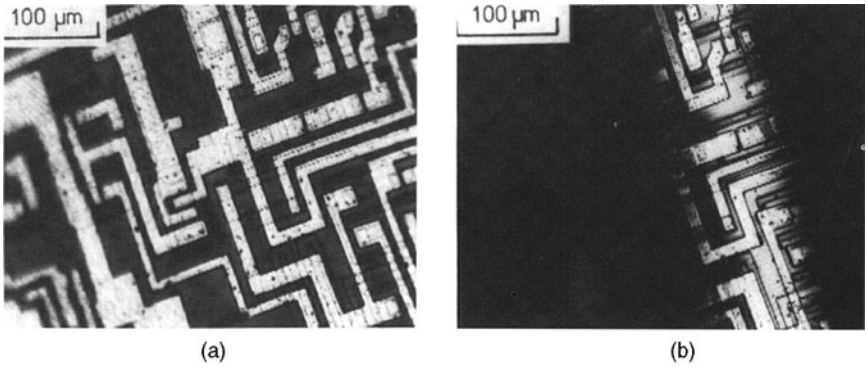


Figure 4.5 A tilted microcircuit imaged with (a) conventional optics and (b) confocal optics, to show the removal of out-of-focus light with confocal systems, taken from Wilson.[6]

4.4, means that an object positioned a few microns above or below the focal plane will appear larger and blurred. However, in a confocal microscope, the short cylindrical PSF means that, as an object is moved out of the focal plane, the intensity is reduced, without blurring, until the object vanishes from the image. The dramatic improvement that results from the different PSFs can be seen in Fig. 4.5, where a tilted microcircuit is viewed with conventional and confocal optics.

In practice, a microscopist will frequently measure the confocal microscope's axial resolution in order to assess the alignment and positioning of the microscope optics. Measurement of the axial resolution by the Rayleigh criterion is clearly quite difficult to achieve in practice, requiring a test specimen that has a number of axially oriented, fluorescent bead pairs separated by known distances. An alternative method is to measure the variation in reflected light intensity as a mirrored slide is passed through the plane of focus. The intensity of a confocal image as a mirror is passed through the plane of focus is shown in Fig. 4.6. The full width half maximum ($\Delta z_{1/2}$) is a convenient parameter for characterising this axial response, giving the 'thickness' of the optical section.

4.2 Modern CLSMs

It was not until a few decades after its invention (and after the original patent had expired!) that confocal microscopy started to make its real impact on research work. In the late-1980s interest in the technique started to accelerate and a whole raft of commercial systems appeared on the market. Most of these systems have a base unit, which contains the laser light source, filters and scanning optics. The scanning output is passed into a conventional microscope (to make use of the existing objective lens and stage mounts).

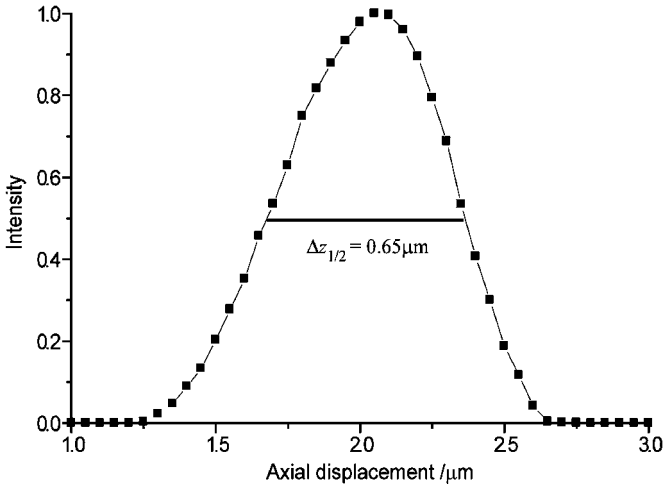


Figure 4.6 A typical axial response for a confocal microscope, taken by measuring the reflected light image of a mirror as it was passed through the focal plane. The full width half maximum, $\Delta z_{1/2}$ of this particular response function is 0.65 μm.

Until recently, the emphasis had been on the 'hands-on' control of the confocal instrumental parameters by the user, but modern instruments have the capability to be controlled fully by a host microcomputer (or workstation). The greatest challenge for the developer of computer-assisted software is to 'remove the microscopist from the decision making process'. In other words, to make the software controlling the operation 'adaptive' – capable of altering the instrumental parameters as a result of the current quality of the recorded images. As seen in Chapters 2 and 3, computer-assisted microscopy has many potential advantages and these CLSM systems are now positioned to make great strides in the materials science research area. An overview of the two commercial confocal microscopes that have been used in the authors' research at Leeds is given in the following subsections.

4.2.1 Biorad MRC 500 series

A highly schematic ray diagram and an overview of the optical path in a typical commercial instrument, the Biorad MRC500, is shown in Fig. 4.7. In this confocal microscope, the XY image frame was built up pixel by pixel by a point-scanning process. The detector was a sensitive, photomultiplier tube located behind the confocal aperture and it received light from a small point in the sample, illuminated by the incoming laser light. The focussed laser light was directed onto the sample by scanning galvanometer-driven mirrors in the optical path. Hence, the time taken to build up the XY frame was usually a second or a few seconds.

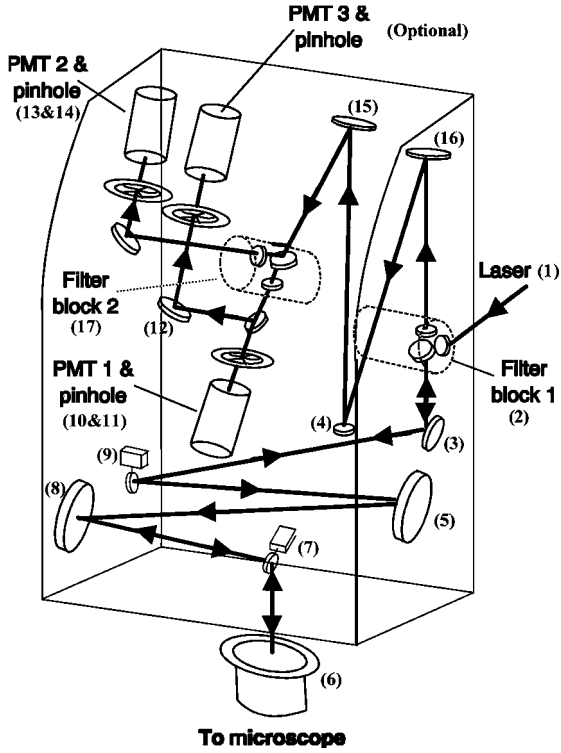


Figure 4.7 Schematic of the folded optics in the Biorad MRC 500 confocal microscope.

As with any commercial CLSM system, there were a number of control functions which had to be mastered in order to optimise the image quality. For example, in the Biorad MRC 500, the main manual controls were:

- confocal aperture size (which influenced the effective thickness of the optical sections, i.e. a small size would produce a thinner optical section)
- electronic gain and dark level (which influenced the image contrast and the brightness achievable)
- neutral density filter selection (which controlled the laser power incident on the sample)
- user selectable filters (so that the user could view in one operational mode and easily switch to another mode before scanning the same sample region). The optical path was complicated by the need to insert these special filters, which selected either the illuminating laser wavelength (the reflection mode of operation) or wavelengths greater than the exciting laser wavelength (the fluorescence mode of operation).

Another potential variable was the type of laser sub-system, which could be incorporated. Longer wavelength laser lines are attractive for materials science because they should penetrate deeper into a sample than shorter wavelengths (the dominant absorption mechanism being Rayleigh scattering, which is proportional to $1/\lambda^4$). However, near-infrared laser systems are more costly and generate lower light intensities than the common argon-ion lasers. In materials science, although higher power lasers can achieve greater depth of penetration for reflection mode studies, they may be inefficient for fluorescence mode applications because of the 'photobleaching' phenomenon (see Section 4.5.1).

It might be thought that the setting of the control parameters above was obvious; make the aperture as small as possible, increase the electronic gain, increase the laser power and see which operational mode gave the better image. However, in practice these parameters were 'coupled' in that the setting of one influenced one or more of the other parameters. Hence the user was involved in an iterative process in order to optimise the CLSM settings for a particular application.

The main drawback with this otherwise excellent instrument was its inability to allow the user to perform semi-automated image data collection. For example, the manual controls had to be pre-set before the start of a run. They could not then be modified to optimise the image acquisition at different depths and over large volumes (which are desirable for most materials science applications). It was clear that another CLSM system would be required with the necessary computer interfaces to replace those Biorad manual control options if the technique were to make significant contributions to materials research.

4.2.2 Noran Odyssey CLSM

The optical schematic diagram of the Odyssey CLSM is shown in Fig. 4.8(a) and also the linkage between different hardware components of the Leeds system is shown in Fig. 4.8(b). The most significant difference between this microscope and the Biorad MRC described in the previous section is the optics used to scan the illumination spot within the focal plane. The Odyssey combines a galvanometer-driven mirror for the slow vertical scanning of the image and an acousto-optical deflector (AOD) for the faster horizontal line scans. The AOD is a solid state device, which consists of a glass substrate with an array of piezoelectric transducers bonded to one end. When driven with a sweeping RF frequency, these create acoustic waves that form a diffraction grating within the crystal. The sweeping frequency varies the pitch of the optical grating and hence the deflection angle of incident light. The AOD has negligible inertia, which enables the device to produce rapid sawtooth raster scans with almost instantaneous fly-back. Hence, this beam scanning system enables the collection of an image every 33 milliseconds, providing the microscopist with a video rate image of the current optical section within the sample. The AOD offers further

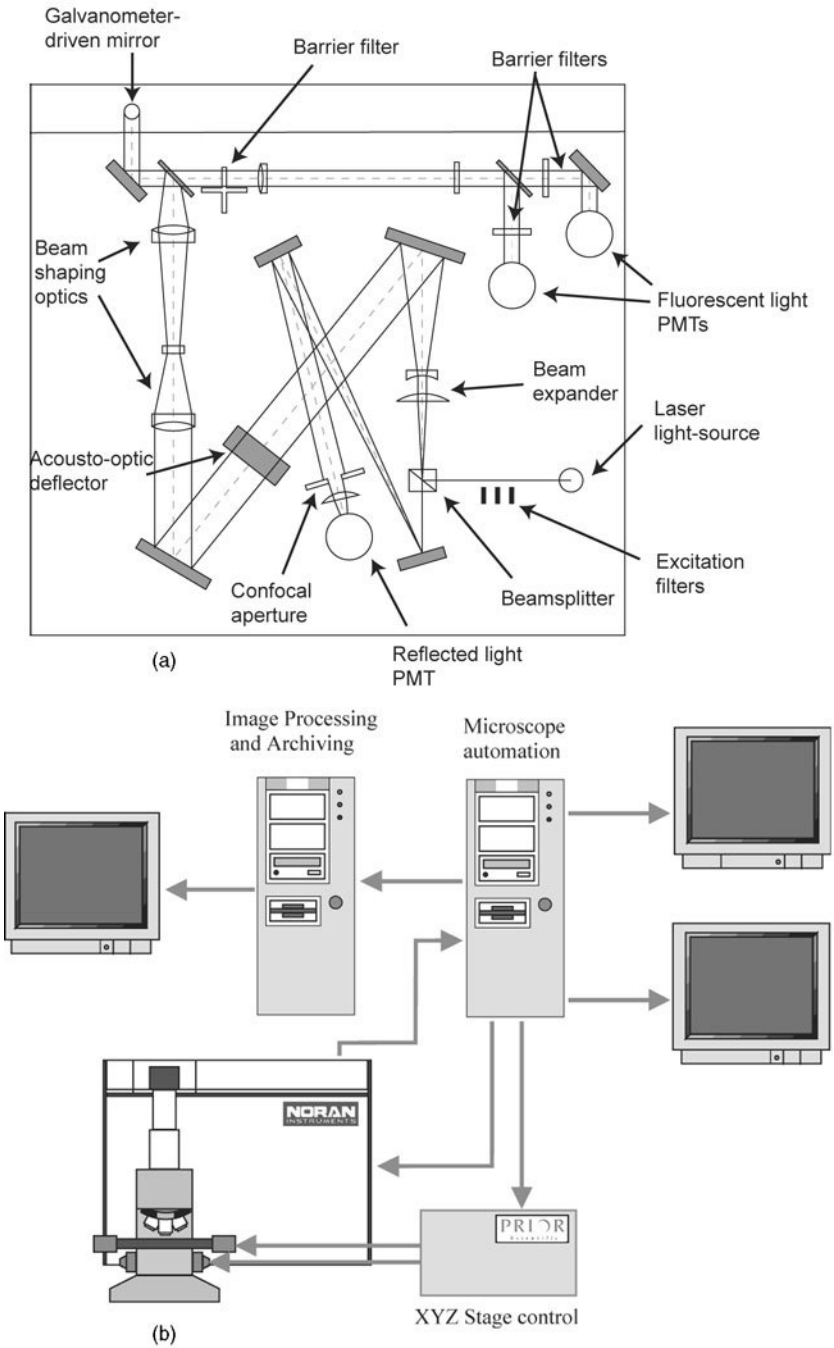


Figure 4.8 (a) Optical schematic of the Noran Odyssey CLSM. (b) The complete CLSM system at Leeds with Odyssey and Prior XYZ stage.

advantages to the microscopist because its diffraction efficiency can be controlled by varying the power input to the AOD, providing a convenient way of controlling illumination intensity and fast shuttering. In addition, the sweep rate of the AOD can be controlled to give continuous pan and zoom.

In the reflection mode of operation, the reflected light follows the same path as the illuminating light beam. The reflected light is first de-scanned vertically by the galvanometer mirror, then horizontally by the AOD and, finally, it is focussed through an adjustable iris onto a photomultiplier tube. In the fluorescence mode of operation, the emitted fluorescence from the sample will be composed of a number of different wavelengths. The deflection angle of the AOD is wavelength dependent, therefore it is not possible to de-scan the emitted light using this device. In the Odyssey CLSM, a separate light path has been added for fluorescence mode operation, where the beam (which is still oscillating linearly) is projected onto a slit rather than a pinhole in front of the fluorescence photomultiplier tube. This mode of imaging, consisting of point illumination and slit detection, results in a radially asymmetric PSF and a slight reduction in axial resolution.

The majority of the manually-adjustable, Biorad MRC 500 controls listed in Section 4.2.1, are controlled and adjusted via a host computer in the Odyssey microscope. The computer integration of the hardware components of the Leeds system is shown in Fig. 4.8(b). This has enabled the development of complex scanning routines where decisions can be made by the computer, and the Odyssey optics modified, based on the current image observed during program execution. The system has proved to be indispensable for the successful completion of much of our work, as illustrated by the case studies at the end of this chapter.

4.3 Optical sectioning capability of the CLSM

The stage scanning confocal microscope which Minsky described in his original patent offered many advantages over conventional microscopes. However, in more modern, commercial confocal microscopes, almost all of these features have been lost, except for one – the optical sectioning capability. Clearly, it is this optical sectioning property which has received the most interest, and it has become the *raison d'être* for the modern confocal microscope. In this section, optical sectioning will be considered in more detail and the practical problems which restrict the 'thickness' of the optical section and the maximum depth at which optical sections can be acquired will be studied for 'real' samples. The following will be considered:

- confocal aperture size
- objective lens NA
- refractive index mismatch
- light attenuation by the sample.

4.3.1 Effect of the aperture control

The rate at which the detected intensity falls off, as a reflective sample is moved axially, demonstrates the sensitivity of the microscope's optical sectioning. This property is clearly a function of both the microscope's optics and the sample itself. By considering an idealised object, a thin planar reflector and a point detector, it is possible to deduce the theoretical optical sectioning limit. The optical sectioning ability of a confocal microscope is normally characterised by the full-width half-maximum, $\Delta z_{1/2}$, of the axial response, as was shown in Fig. 4.6. With an idealised sample and detector, this parameter is dependent on the objective lens NA, the laser wavelength, λ, and the immersion medium refractive index, n_i, as follows:[12]

$$\Delta z_{1/2} = \frac{0.43\lambda}{\left\{ n_i - n_i \cos\left[\sin^{-1}\left(\frac{\mathrm{NA}}{n_i} \right) \right] \right\}} \qquad [4.5]$$

The function described by equation [4.5] is shown in Fig. 4.9 for both air and oil immersion objective lenses at two different wavelengths.

Modern confocal microscopes normally have a variable aperture positioned in front of the detector, which allows the operator to adjust the amount of light reaching the detector. Clearly, the size of this aperture has an effect on the sectioning strength of the microscope. The smaller the aperture, the closer the sectioning strength comes to the limit of an ideal detector given by equation [4.5], whereas, in the limit of a very large aperture, the sectioning property disappears altogether. However, the aperture size also affects the amount of light which reaches the detector, thus making very small aperture sizes impractical. Figure 4.10(a) shows the axial response (together with their respective values of

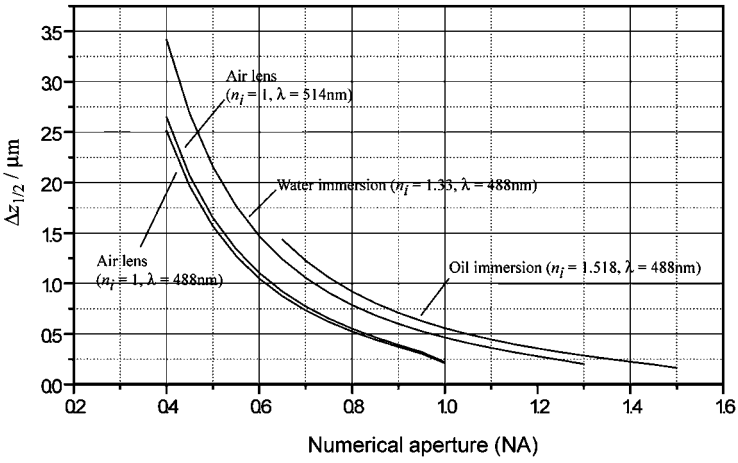


Figure 4.9 The axial resolution is characterised by $\Delta z_{1/2}$, which is plotted in terms of the numerical aperture of three popular types of objective lens.

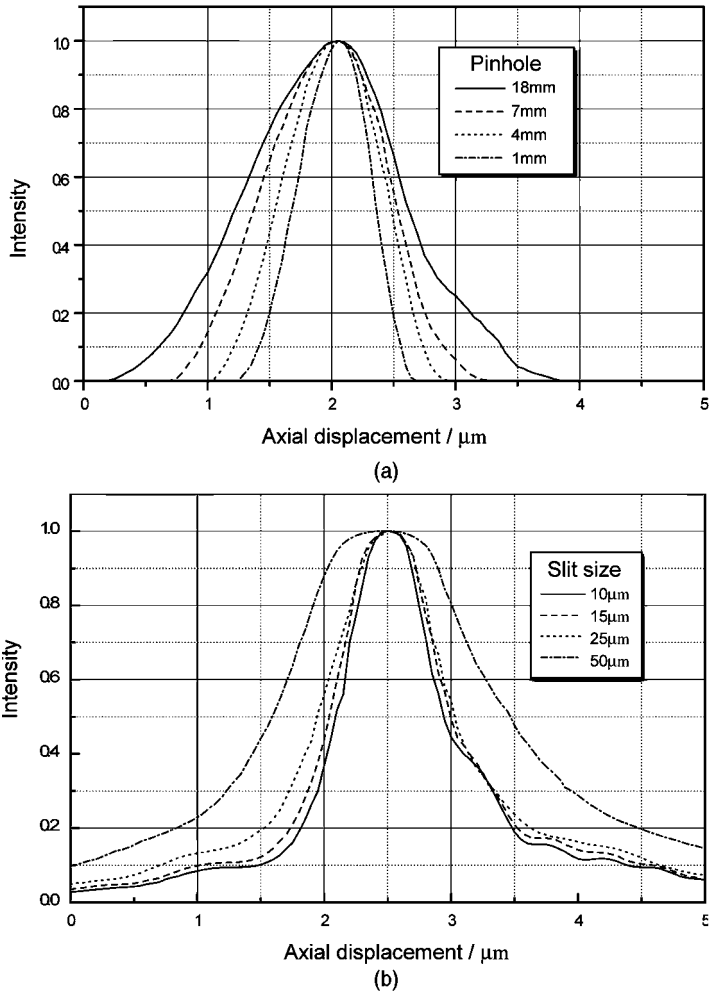


Figure 4.10 (a) The axial response of the Noran Odyssey CLSM with various aperture sizes and (b) the axial response of the fluorescent light path with various slit sizes.

$\Delta z_{1/2}$) of the Noran Odyssey at various aperture sizes. The aperture sizes used with the Odyssey are somewhat different to most confocal microscopes due to the beam expanding optics placed just in front of the reflected light detector. These optics allow an adjustable iris, which gives aperture sizes in the range 1–18 mm to be used as the 'pinhole'. However, most confocal microscopes use apertures within the range 5–150 μm. For comparison, Fig. 4.10(b) shows the measured axial response for the fluorescent light path of the Odyssey microscope. Note the slight loss in axial resolution introduced by replacing the pinhole with a slit.

4.3.2 Refractive index and the apparent depth problem

With a conventional light microscope, images are normally acquired on, or very near, the surface of a sample. If images are to be acquired from within the body of a thick sample, physical sectioning is required. Therefore, if the microscopist selects a suitable objective lens for the magnification and field of view required, selects the correct immersion medium for this objective lens (and cover slip if required) then a good quality image should be observed. However, with a confocal microscope, sections are acquired optically and are often at some significant depth within the sample.

The path followed by the illuminating and reflected light through the sample has a significant effect on the observed image quality. If the sample has a good refractive index match with the objective lens immersion medium, the light cone will converge as intended. On the other hand, if the sample has a significant difference in refractive index, spherical aberrations will occur. Figure 4.11 illustrates the occurrence of spherical aberrations when an oil immersion objective is used to image deep within a sample that has a significantly different refractive index. The marginal rays no longer converge with those rays closer to the optical axis, therefore the light is no longer focussed to a single point, resulting in a degradation of the image quality. If possible, samples should be tailored to match the refractive index of the objective lenses available, or vice versa. However, in practice it is rarely possible to modify the refractive index of the sample and there is a limit to the range of suitable objective lenses.

If there is a refractive index difference between the sample and the immersion medium, this will result in optical sections where objects are axially blurred and there will be a shift in the perceived depth of focus. A number of research groups have tackled the problem of extracting accurate data from optical sections affected by spherical aberrations. Visser *et al.*[13] proposed a simple ray tracing model for predicting the apparent shift in focal plane. In their model, when the microscope stage moves vertically by an amount Δs, the movement of the focal plane is actually Δz, where the equation linking the two quantities is:

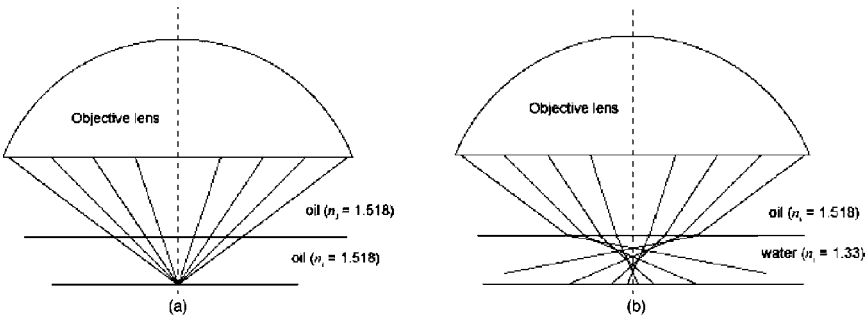


Figure 4.11 (a) An oil immersion objective lens imaging without aberration and (b) with spherical aberration due to refractive index mismatch.

$$\Delta z = \frac{\tan(\sin^{-1}(\mathrm{NA}/n_i)}{\tan(\sin^{-1}(\mathrm{NA}/n_s)} . \Delta s \qquad [4.6]$$

where n_i and n_s are the refractive indices of the immersion medium and sample respectively.

Note that Visser's correction is based on the following assumptions:

- only the marginal rays in the objective acceptance cone are used to determine the focal point
- the errors, including spherical aberrations in the microscope's optical system, or the interface between immersion medium and sample, are ignored
- the immersion medium and the sample are both considered to be homogeneous.

However, Carlsson demonstrated, by restriction of the objective aperture, that the paraxial rays have a greater contribution to the image.[14] Therefore, these results cast some doubt on assumption (a) in Visser's model and it is probably best to treat Δz as given by equation [4.6] as an upper bound. More recently, research has been concentrated on formulating a complete description of refractive index mismatch-induced, spherical aberrations by integrating over the entire incident light cone.[15] Therefore, it is possible to reconstruct the shape and intensity of the aberrated PSF at different depths within a mismatched medium, as seen in Fig. 4.12. The results of this analysis show that, as the lens is focussed deeper within a refractive mismatched sample:

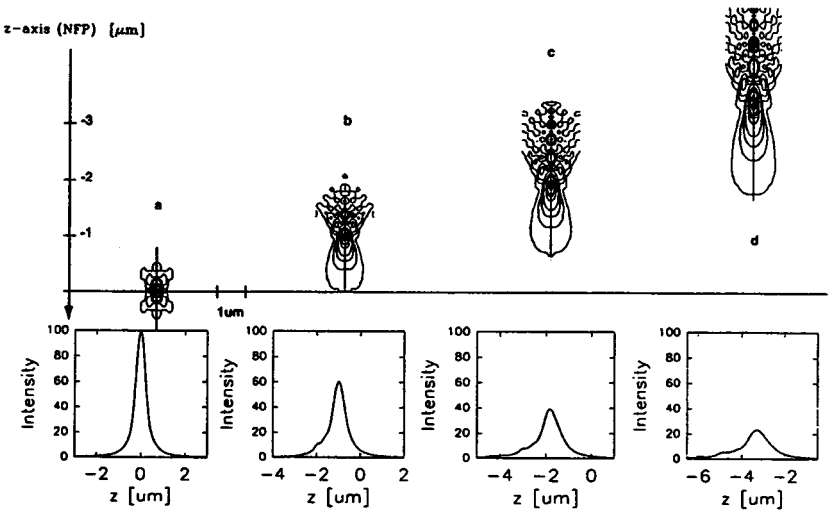


Figure 4.12 Contour plots of the aberrations in the confocal PSF as an oil immersion lens is focussed through a sample with the same refractive index as water (from Hell *et al.*[15]).

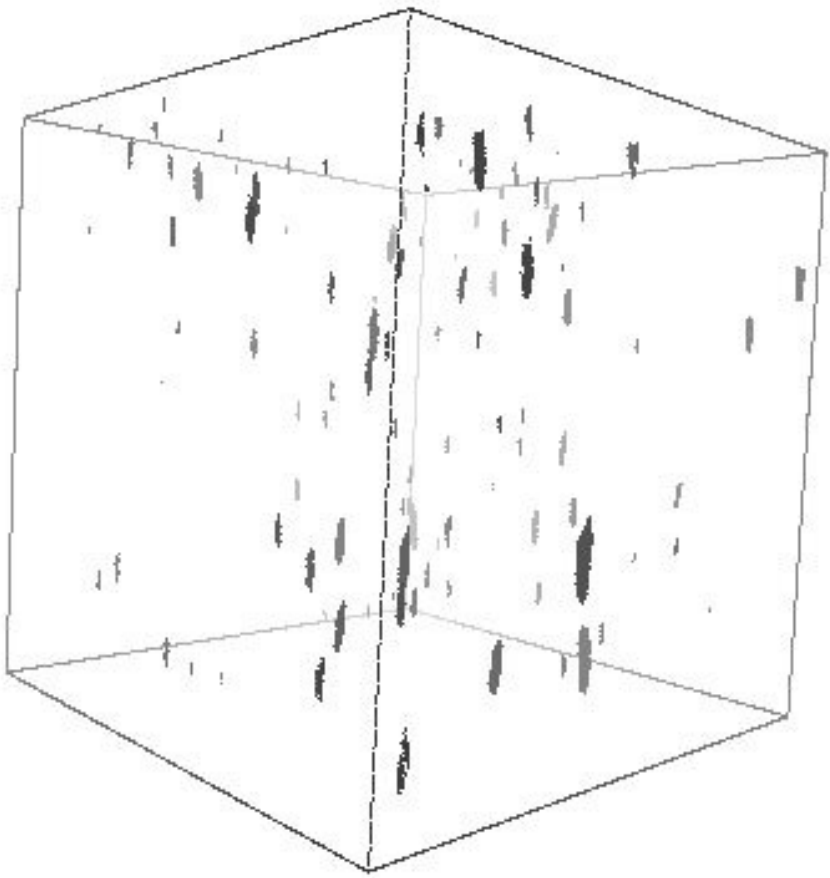


Figure 4.13 A number of particulates embedded within a clear polymer film. The top and bottom surfaces of the film are both visible. Although the number of particulates appears to decrease with depth, this is in fact an artefact due to the reduction in peak intensity of the PSF with increased depth. The volume shown measures 100 μm $\times$ 100 μm $\times$ 100 μm in XYZ.

- there is a significant loss in axial resolution and a less significant loss in lateral resolution
- the intensity of the PSF peak decreases
- the location of the peak is shifted axially.

The implications of these effects are demonstrated in Fig. 4.13, where particulates within a thick Melinar disk (refractive index of 1.6) have been imaged with a low magnification, air objective lens. The decrease in the PSF peak intensity results in an apparent decrease in particle number with depth. However, the collection and reconstruction of volume datasets using different

gain levels at different depths demonstrated in Section 4.6 allows the collection of more reliable datasets.

4.3.3 Maximum useable depth

The depth of penetration is actually the most significant factor when considering using a confocal microscope in research work – coupled with the spatial resolution that is required. It has already been seen that, to improve the optical sectioning strength, the objective lenses must be characterised by a large numerical aperture (NA) and oil immersion lenses have the highest NA values. Sadly, these lenses also have the shortest available working distances (the distance from the end of the objective lens to the point of focus), so there is a natural trade-off between spatial resolution and the ultimate depth of penetration.

As can be seen in Table 4.1, the oil immersion lenses have working distances around 200 μm whereas the air lenses have working distances of a few millimetres. In practice, however, this penetration will rarely be achieved because of refractive index mismatching, the attenuation of light within the sample and the quality of the image processing software. Each of these issues will be addressed in the following sections.

4.3.4 Attenuation of light within samples

An ideal sample for achieving maximum depth penetration (up to the working distance of the objective) and optical sectioning strength would have its refractive index matched to the objective lens refractive index throughout its volume. Unfortunately, such a featureless sample would not be very interesting to study! 'Real' samples tend to be heterogeneous and include objects which reflect light or which fluoresce strongly. However, these objects will also scatter light and cast shadows, effectively attenuating the illumination at greater depths.

As an example, Fig. 4.14 shows fluorescence XZ sections (see Section 4.5.3) of two different glass fibre-reinforced composites. In both samples, the matrix auto-fluoresces (i.e. the matrix has not been stained by a fluorescent dye) and

Table 4.1 Working distances of typical CLSM objectives

Manufacturer	Type of objective	Numerical aperture	Magnification	Working distance
Nikon	planapo, oil	1.4	×60	220 μm
Zeiss	planapo, oil	1.4	×100	100 μm
Zeiss	achroplan, water	0.9	×63	1.45 mm
Zeiss	eriplan, air	0.75	×100	3 mm

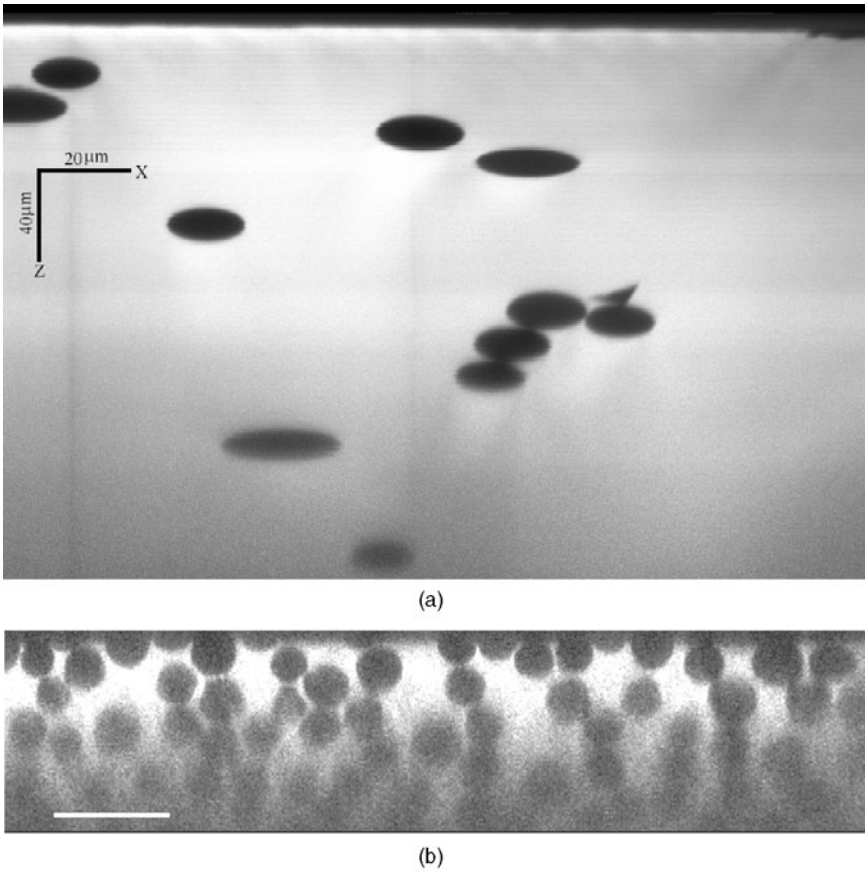


Figure 4.14 A comparison of the maximum depth penetration for (a) a low packing fraction carbon fibre-reinforced composite (the depth achievable corresponded to the full working depth of the ×60, NA1.4, oil immersion, objective lens) and (b) a high packing fraction, glass fibre-reinforced composite illustrates the problem of depth penetration below 40 μm. (Three XZ sections have been overlapped to show an area 220 μm × 50 μm.) Both images were recorded in the fluorescence mode of operation.

provides the contrast between fibre and matrix in these images. With the low fibre packing fraction sample, useable images have been acquired over the full working distance of the objective lens. However, in the higher packing fraction sample, the light is severely attenuated due to light reflected at the fibre-matrix interface casting shadows. There is a good refractive index match between the glass-reinforced composite and the immersion medium (refractive index of 1.5), and therefore any attenuation due to spherical aberrations is minimal. Figure 4.15 shows the exponential nature of the loss in signal intensity with depth.

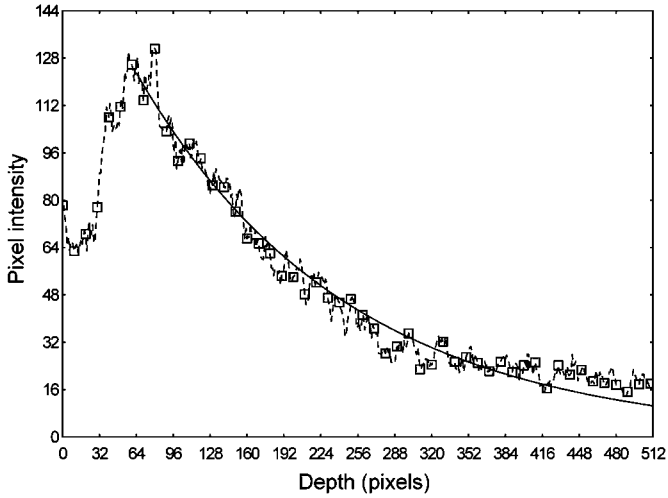


Figure 4.15 A plot of the pixel fluorescence intensity as a function of depth showing the exponential nature of the attenuation of the fluorescence signal.

4.3.5 Interpretation of data

The sample volume which can be analysed ultimately depends on the maximum useable depth, and this parameter depends on many factors, all of which affect the specimen's optical transparency. For example, the maximum useable depth depends on the density and the clustering of scattering centres within a sample and whether or not different parts of the sample fluoresce. But most importantly of all, it is the sophistication of the image processing routines which ultimately determines the sample volume that can be reconstructed within reasonable timescales.

It is therefore difficult to compare different CLSM systems and, ultimately, the sample volume reconstructed depends on whether or not the reconstruction is to be aided by operator intervention or by the automation of as many processes as possible. For example, one might achieve greater depth of penetration into the sample by using an interactive program which allows the operator to pick out objects within the XZ image fields, but such a process would be time-consuming and tedious. Better to make the image processing as reliable as possible for picking out objects – even if to smaller useable depths – and to cover these depths over larger sample areas.

Although the image data acquisition may be fully automated, the image data reduction may have to be followed by a manual 'veto' operation. The manual veto checks the sense of the image data (because of the considerable difficulties associated with detecting well-known geometrical objects within fuzzy, noisy image frames).

Note that a number of groups are using a lot of computer power to improve the classification of objects well below the surface. They achieve the classification by deconvolving the CLSM images or allowing for ray tracing through upper layers of the sample material (e.g. Visser *et al.*[16] and Kriete[17]). Ultimately one must decide the trade-off between image processing times and sample scanning times.

4.4 Calibration issues

The accurate calibration of the microscope's field of view is essential if it is going to be used for quantitative measurement. In Chapter 1, an overview was given of the historical evolution of the microscope optics which are used in today's instruments. Many of these innovations improved the quality of the image observed by the microscopist ensuring that it is planar, uniform in intensity and free from distortion. However, the route that the image takes from the microscope to its final digital form can be complex and distortions may arise. Furthermore, with laser scanning confocal microscopes, the linearity and intensity of the image is dependent not only on the integrity of the objective lens (and other microscope optics) but also on the scanning optics. Both the commercial confocal microscopes described in the previous section were found to have significant distortions because of their scanning optics. The steps that were undertaken to correct the images from both microscopes are described.

4.4.1 Shading correction

As mentioned previously, the beam scanning within the Noran Odyssey is performed by a combination of galvanometer-driven mirror, in the Y direction, and an AOD in the X direction. Figure 4.16 shows an image of a mirror acquired in reflection mode, demonstrating that these optics produce an illumination/collection intensity that is non-uniform across the field of view, with the majority of the non-uniformity being attributed to the AOD. This uneven illumination is less apparent in fluorescence mode where the light emitted by the sample follows a different path to the illuminating light.

A shading correction can be achieved digitally by the collection of a reference image, as described in Chapter 2. If the same field of view is used throughout a study (i.e. electronic pan and zoom are not applied), the reference image need only be acquired once. In a confocal microscope, the collection of such a reference image is complicated by the axial response of the microscope. If the mirror is not perfectly orthogonal to the optical axis of the microscope, the image will intersect the axial response function at different points across the field of view, as illustrated in Fig. 4.17. This can be avoided by collecting auto-focus images, where the sample is moved through the focal plane and the maximum intensity at each image pixel is recorded (described later in Section

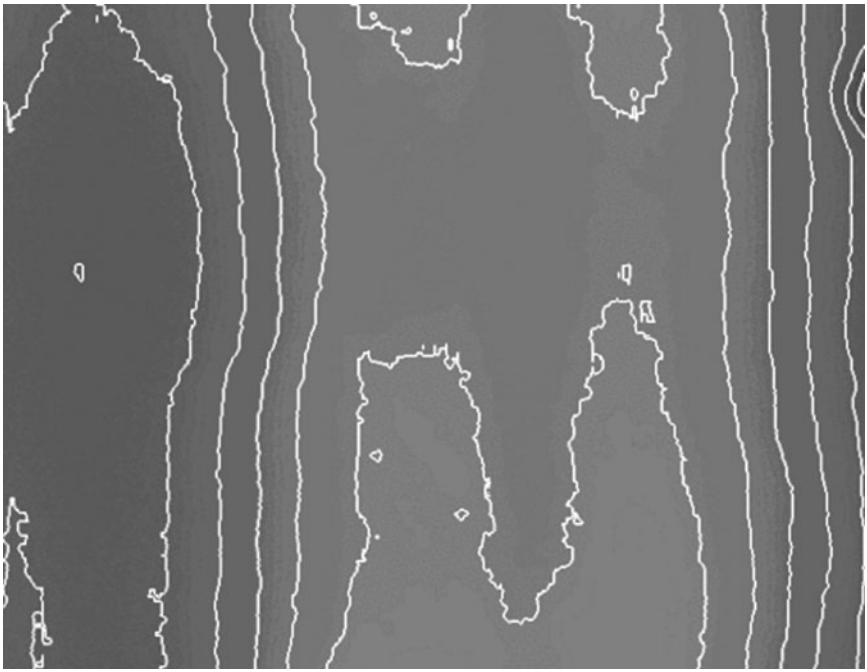


Figure 4.16 A reflection mode image of a mirror clearly demonstrates the non-uniform image gain over the field of view. Intensity contours are plotted for each intensity step of 10.

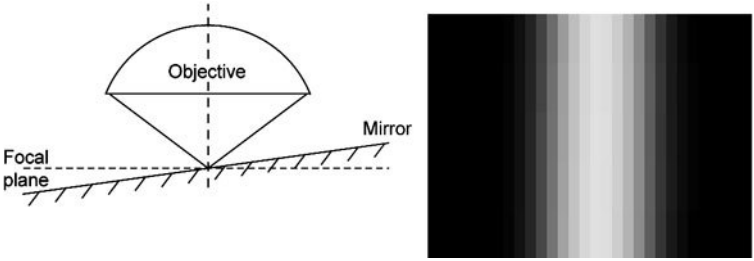


Figure 4.17 When an inclined mirror is viewed in reflection mode by a confocal microscope, as shown in (a), the resulting image shown in (b) is the axial response function.

4.5.4). In fluorescence mode, an alternative test specimen for calibrating the microscope is a thick block of fluorescent material, where the isotropic emission of light removes the need for auto-focus imaging.[18]

In Chapter 2, the following equation was introduced for the shading correction, where $I_o(x,y)$ is the original image intensity, $I_c(x,y)$ is the corrected intensity, $B_r(x,y)$ is the reference background image, $B_e(x,y)$ is an image

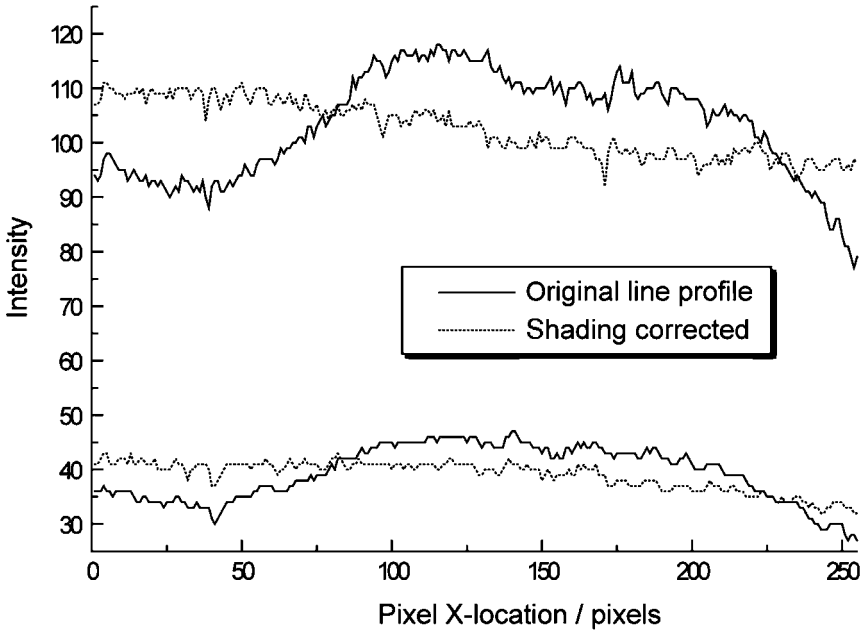


Figure 4.18 Line profiles through an image of a mirror at two different gain settings. Dotted lines show the line profile after background shading correction.

acquired when the illumination intensity is zero and $< I_o >$ is the mean intensity of the original image.

$$I_c(x,y) = \frac{I_o(x,y) - B_e(x,y)}{B_r(x,y) - B_e(x,y)} < I_o > \qquad [4.7]$$

Figure 4.18 demonstrates the effect of background correction using the above equation, by comparing line profiles from the original and corrected image. Notice that the corrected images do not have a constant intensity, as would be expected from an image of a mirror. This is because the mirror is not perfectly orthogonal to the microscope's optical axis, and hence demonstrates the importance of acquiring auto-focus images. The angle of slope of the mirror can be calculated by comparing the intensity variations across this line profile with the axial response function in Fig. 4.10(b). In this case, a ten per cent variation in intensity across a distance of 100 μm within the image leads to the conclusion that the mirror is sloped at approximately 0.1° to the XY (focal) plane of the microscope.

4.4.2 Geometric correction

With stage scanning confocal microscopes, the slow scanning mechanism enables accurate speed control while executing the raster scan pattern. With

beam scanning confocal microscopes and, to a greater extent, real-time beam scanning microscopes, the scan rate may be non-uniform, introducing distortions into the image. The severity of the distortion depends on the optical devices used to perform the beam scanning. The galvanometer mirror that is used in the majority of commercial confocal microscopes, if properly maintained, should not introduce a significant distortion to the image.[19] However, even the smallest distortions result in a mismatch between neighbouring frames when employing the large area scanning routine described in Chapter 3.

The effect of image distortions can be assessed by measuring the size of a known feature at different positions within the field of view; a graticule with a regular square grid pattern (of known dimensions) is a suitable test specimen. The microscopist can measure the apparent size of the graticule squares at various positions within the field of view. Then, by dividing the dimensions (in pixels) of the highlighted square by its known dimensions (in microns), the magnification, at this particular point within the field of view, is computed. An interactive routine has been used to measure the stretching of the image produced by the Noran Odyssey and the results are shown in Fig. 4.19 as the variation in magnification in the X and Y directions. This variation has a significant effect on line measurements, with objects appearing six per cent longer on the right side of the screen, when compared to the same object imaged on the left side of the screen.

In Chapter 2 the correction (and application) of geometric distortions was described as a smooth transformation of an image $t(x,y)$, from a source image $s(x,y)$ via a pair of functions $i(x,y)$ and $j(x,y)$.

$$t(x,y) = s(i(x,y), j(x,y)) \qquad [4.8]$$

The observed distortions in this system were found to be approximately linear in both X and Y without correlation between the two, which results in a simpler description of the image distortion:

$$t(x,y) = s(i(x), j(y)) \qquad [4.9]$$

By fitting straight lines to the data, the following simple description of the image distortion was obtained:

$$\begin{aligned} i(x) &= 0.95x(1.015 - 5.4 \times 10^{-5}x) \\ j(y) &= 0.96y(0.99 + 8.8 \times 10^{-5}y) \end{aligned} \qquad [4.10]$$

In practice, a single pixel on the distorted image, *s*, will not always yield integer values for its location when transformed onto the target image, *t*. For example, the pixel at the location *t*(150, 240) takes the intensity of the pixel at the location *s*(146.3, 242.7) in the original image. The transformed location could be truncated to the nearest integer value (or to the closest in a Cartesian sense); however, this results in unpleasant aliasing effects. A better method is to interpolate a new intensity value from the four nearest neighbours as illustrated

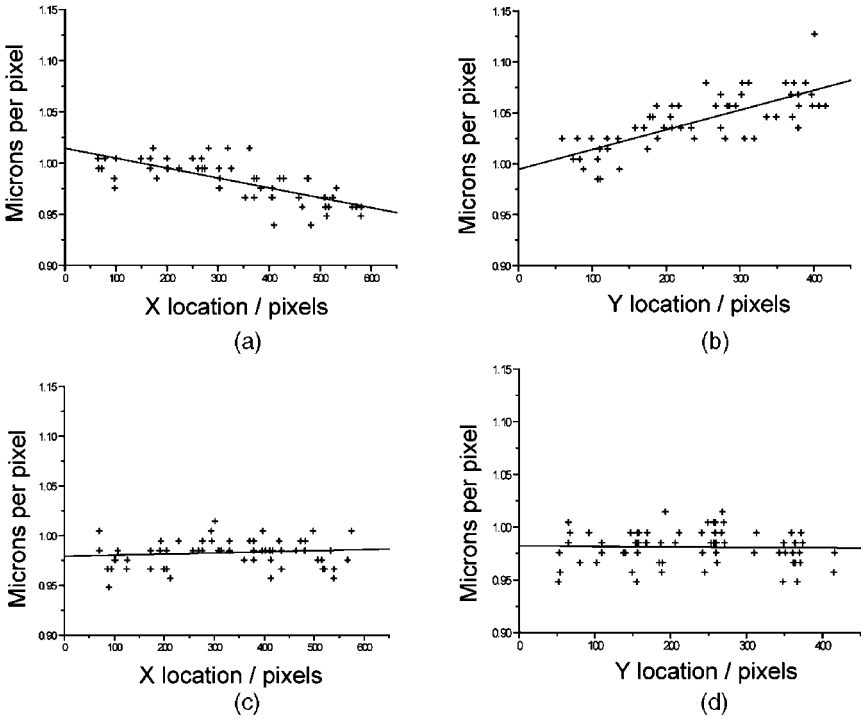
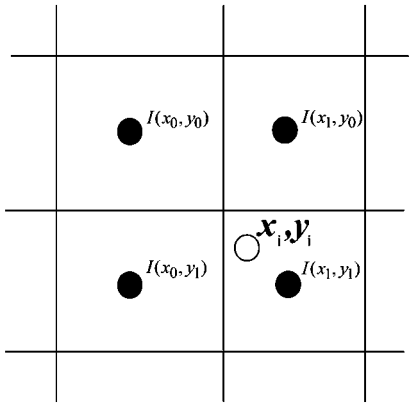


Figure 4.19 (a) and (b) The variation in magnification across the field of view in the X and Y directions respectively. (c) and (d) The variation in magnification across the field of view, for a stretch corrected image, in the X and Y directions respectively. These data were acquired with a ×10, NA 0.4, air objective lens.

in Fig. 4.20. The pan and zoom facility, which is available in most modern CLSMs, is performed by changing the scan range of the galvanometer-driven mirrors (and the AOD in the Odyssey). This will almost certainly result in a change in the way the image is distorted and the calibration procedure must be repeated.

4.4.3 Z calibration

The issue of the Z calibration of a CLSM system is more problematical and much work has been, and is being, invested in manufacturing a suitable standard with which to assess the Z calibration, see for example Cogswell.[20] For the same reasons as discussed in Chapter 3, one cannot assume that, by sending a fixed number of pulses to a stepper motor-driven Z movement, one obtains the same linear movement every time! (Although expensive piezoelectric transducers are now being used to give high precision movement and better linearity of movement.) However, for materials science applications, the issue of the Z



$$J(x_j,y_j) = I(x_0,y_0)(x_1 - x_i)(y_1 - y_i) + I(x_1,y_0)(x_i - x_0)(y_1 - y_i) + I(x_0,y_1)(x_1 - x_i)(y_i - y_0) + I(x_1,y_1)(x_i - x_0)(y_i - y_0)$$

Figure 4.20 A diagram showing the interpolation of pixel intensities from the stretch corrected image to the original image.

calibration is of secondary importance, when compared to the potential problem of the 'apparent depth' of artefacts below the sample surface. This effect is due to refractive index mismatching and has been discussed in more detail in Section 4.3.2.

4.5 Imaging modes

We have seen in Chapter 2 how the optics of a conventional microscope can be modified to enhance different features of the sample being studied (for example, phase contrast imaging). Confocal microscopes offer the microscopist considerable flexibility over how they image samples and the interpretation of the image data by the host computer. The most obvious method of collecting data is to systematically section the material throughout the entire useable volume and to store these data as a volume dataset. However, in practice, there are more effective ways of presenting and storing the sectioned data. A number of novel imaging techniques have been developed that take advantage of the improved depth discrimination of confocal microscopy and some of the more commonly used methods will be described in this section. When combined with the reflection and fluorescence imaging modes, these techniques provide the researcher with a very flexible imaging system.

4.5.1 Reflection and fluorescence modes of operation

As has been mentioned already, there are two main CLSM operational modes. The reflection mode is the term used to denote that the detector receives light at

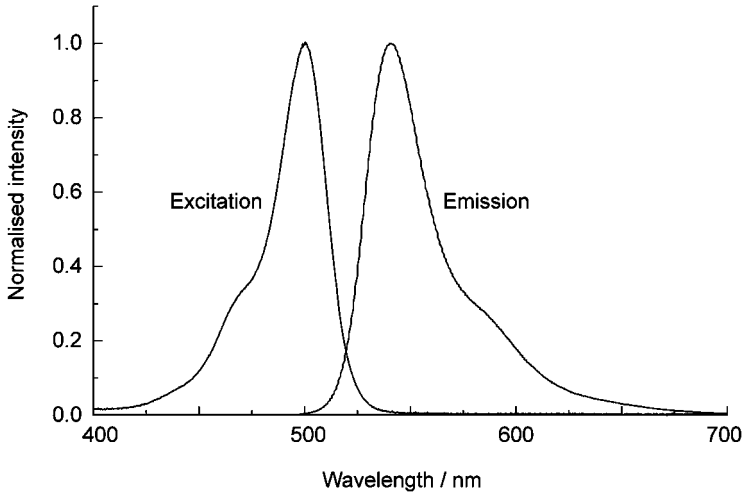


Figure 4.21 A plot of the excitation and emission spectra for fluorescein dye, which allows the microscopist to choose a suitable illumination wavelength and barrier filter set-up.

the same wavelength as the exciting laser radiation. Fluorescence mode is the term used to denote that the detector receives the fluorescence radiation that has been excited by the incident laser light but radiates at longer wavelengths.

Fluorescent dye can be added to a sample in order to reveal features that do not reflect light or highlight areas of interest. Dyes such as fluorescein and rhodamine are used extensively within biological applications; however, their application is less common in materials science. The excitation and emission spectra for fluorescein are shown in Fig. 4.21. From these graphs, it can be seen that a suitable illumination wavelength would be 488 nm and a suitable barrier filter for the resulting fluorescent signal would be a 550 nm long-pass filter. Figure 4.22(a) shows two images within a foam sample, which has been used in an experiment to determine the penetration of particulates in airflows. The particles have been labelled with fluorescein so that they can be identified immediately in fluorescence mode, whereas the structure of the foam is visible in reflection mode. As a result of this work, the particulate penetration at different locations across the face of the foam filter could be obtained, as shown in Fig. 4.22(b).

Some samples may readily fluoresce and an example is shown in Fig. 4.23 where the surface of a polymer composite is imaged in both reflection and fluorescence modes. The polished ends of the fibres reflect light very well, providing contrast with the matrix. In fluorescence mode, the polymer matrix is found to auto-fluoresce, whereas the glass fibre does not, effectively giving the negative of the reflection mode image. Note that with this sample the fluorescence signal was weaker than the reflection signal. Hence, it was

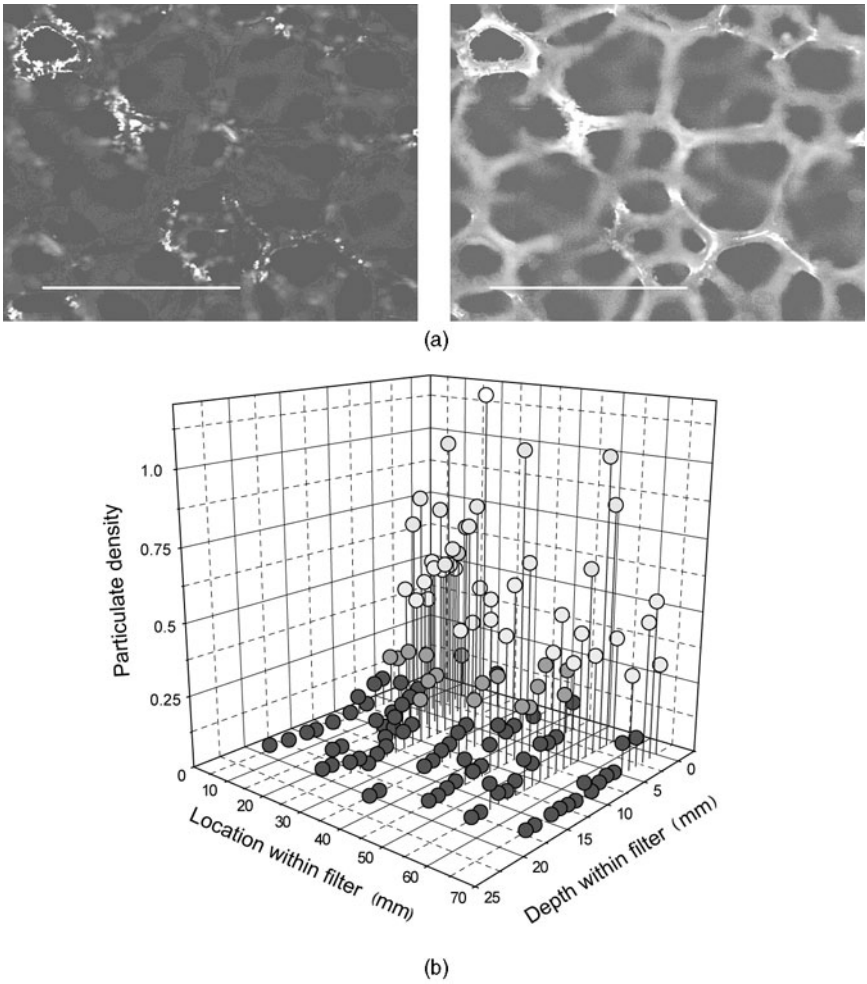


Figure 4.22 (a) A fluorescence mode image and a reflection mode image of a polyurethane foam sample with labelled particulates. The scale bar represents a length of 1 mm. The location and number of the particulates are easily determined from the fluorescence mode image, resulting in (b) a plot of particulate penetration within the foam.

necessary to increase the laser power and integrate over many image frames to improve the signal-to-noise ratio and hence the contrast between fibre and matrix. However, with increased laser power, there will be irreversible chemical changes to fluorescing molecules, which inhibits the fluorescence. This photochemical damage, called photobleaching, is one of the most important and yet least understood aspects of the use of fluorescence microscopy (see Tsien and Waggoner.[21]) The practical effect of photobleaching is to gradually decrease the intensity of the fluorescence signal with time.

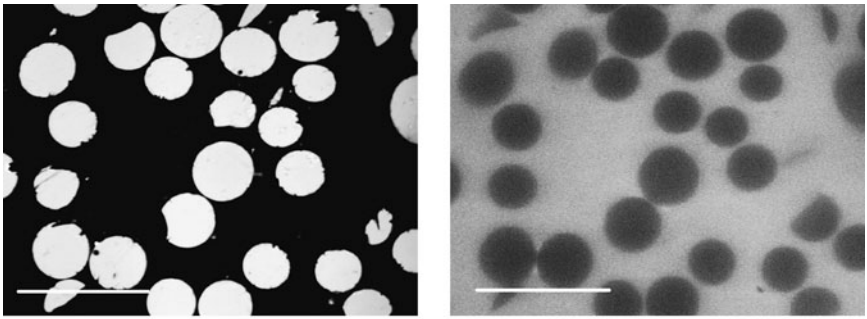


Figure 4.23 (a) Reflection mode image at the surface and (b) a fluorescence mode image near the surface of a glass fibre-reinforced, polymer composite sample. The scale bar represents a length of 20 μm. Note the polishing artefacts (cuts, chips and fragments) in the reflection mode image which are not present in the fluorescence mode image.

Note that all of those sample regions within the light cone during confocal scanning are exposed to large laser powers, even though they are not being imaged. For example, if sixteen optical sections are to be recorded at sixteen different depths within the sample, the sixteenth section will have been photobleached by probably ten to forty per cent by all of the previous fifteen scans! Furthermore, the size of the raster scan executed by the laser beam has a significant effect on the exposure time of the sample. For example, in some CLSM systems, the Y scan can be stopped in order to accelerate the collection of XZ sections. Figure 4.24 shows an XY image where the photobleaching damage confirms the location of an XZ scan acquired previously. Although the photobleaching limits the ultimate signal-to-noise ratio, it can be useful for large area scanning algorithms, providing proof of the location of previously scanned regions!

Some CLSM machines also offer a third mode of operation – the 'photon counting' mode – which can be used to detect very weak signals, as discussed in Art.[22] Although this is a very sensitive technique, images are built up over long time periods (compared to the time needed to create images using other modes of operation).

4.5.2 Volume reconstruction and visualisation

When two-dimensional images are produced, little consideration of visualisation techniques is needed for displaying the data. However, confocal microscopes routinely produce three-dimensional datasets, where a number of optical sections are collected at regular increments, Δz, along the optical axis. With most commercial confocal microscopes the collection of a volume dataset, which typically contains in the region of 100 images, takes no more than a few minutes. Clearly, this is a very rapid technique for collecting vast amounts of

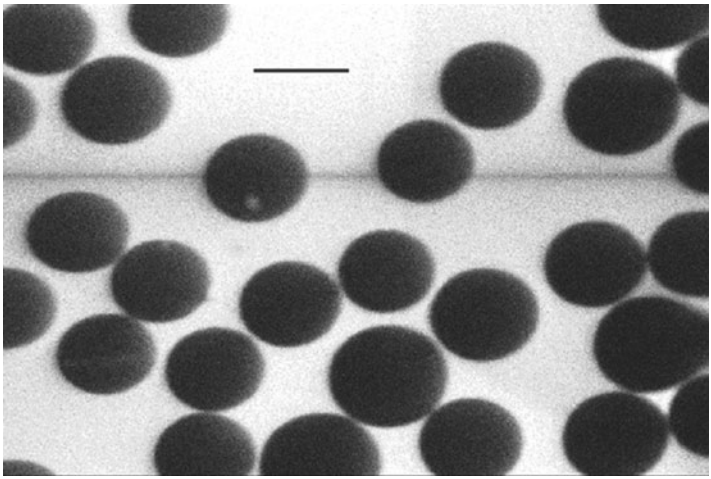


Figure 4.24 An XY fluorescence mode image of a fibre-reinforced polymer composite showing the photobleaching that has resulted from the acquisition of three parallel XZ sections. The scale bar represents a length of 20 μm.

image data. In order to exploit fully the possibilities of higher dimension datasets, many 3D visualisation techniques have been developed and they have proven very popular in recent literature. The aesthetic qualities of the 3D renderings found in recent publications (primarily in the biological sciences), has played no small part in the rise in popularity of confocal microscopy.

In order to render a volume dataset, a virtual screen is positioned with respect to the volume dataset as illustrated in Fig. 4.25. The rendering process involves extending vectors from the virtual screen and locating their intersection with the volume dataset. In this figure, the vectors are all parallel, thus producing an image without perspective; however, if these vectors were divergent the resulting projection would have a sense of depth with perspective. The manipulation of the virtual screen location and orientation allows the microscopist to view the data from any direction, with any orientation.

Two methods of projecting the dataset onto the virtual screen are commonly used:

- *Surface rendering:* the 3D dataset is transformed into a polygonal surface by the selection of a suitable intensity. This iso-surface is converted into an optimised surface mesh using techniques such as the marching cubes[23] or Delauney triangulation[24] algorithms. The colour and intensity can be determined from its orientation with respect to a light source and shading techniques applied.[25] Finally, each polygon is projected onto the virtual screen in turn, as defined by its respective z-location (see Fig. 4.34(b)).
- *Volume rendering:* in the surface rendering technique, rays are traced from the polygon mesh dataset onto the virtual screen. However, in the volume

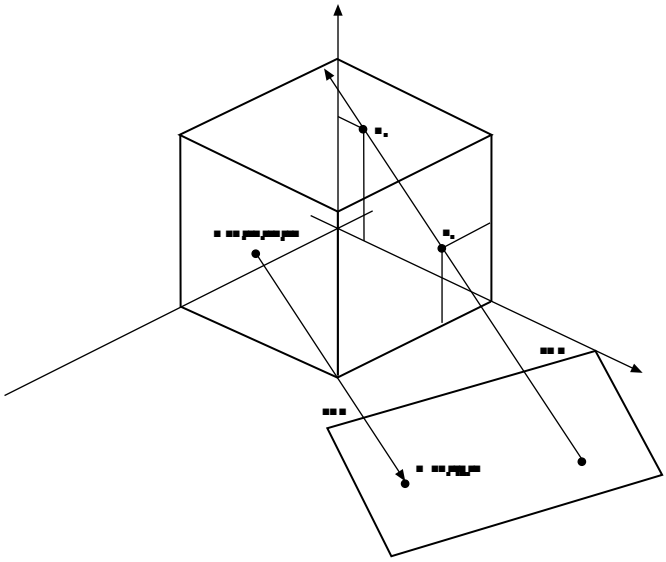


Figure 4.25 A schematic diagram illustrating two methods of projection. (a) Each voxel is considered in turn and a vector is constructed that passes through the voxel and is normal to the virtual screen. The point of intersection on the virtual screen gives the projected position of the voxel. (b) Each pixel on the virtual screen is considered and a vector is constructed that is normal to the virtual screen. The two points at which the vector intersects the square faces of the 3D dataset are located (I_1, I_2). The path joining these points is calculated by a fast line algorithm, hence locating all the voxels that are projected to this point on the virtual screen.

rendering technique, rays are traced in the other direction, that is from the virtual screen towards the dataset. Each ray will intersect a number of voxels (the 3D analogue of pixels) along its path and a blending algorithm that combines the intersected voxels determines the final intensity and colour of each pixel on the virtual screen. This blending, which may be a function of voxel intensity, z-position (with respect to the virtual screen) and some predetermined opacity, determines the appearance of the rendered image.[26] As an illustration of what is possible, Fig. 4.26 shows a volume rendered reconstruction of a fibre-reinforced composite and Fig. 4.27 a volume reconstruction of a polyurethane foam sample.

Of these two rendering methods, volume rendering is the most commonly used visualisation technique for biological specimens due to the 'organic' appearance of images produced by the blending algorithm.[27] Typical samples in materials science, which are not as transparent as biological samples, lend themselves more to the surface rendering technique.

The main disadvantage of these techniques, where entire sample volumes are stored for 3D visualisation after acquisition, is the vast amount of data that will

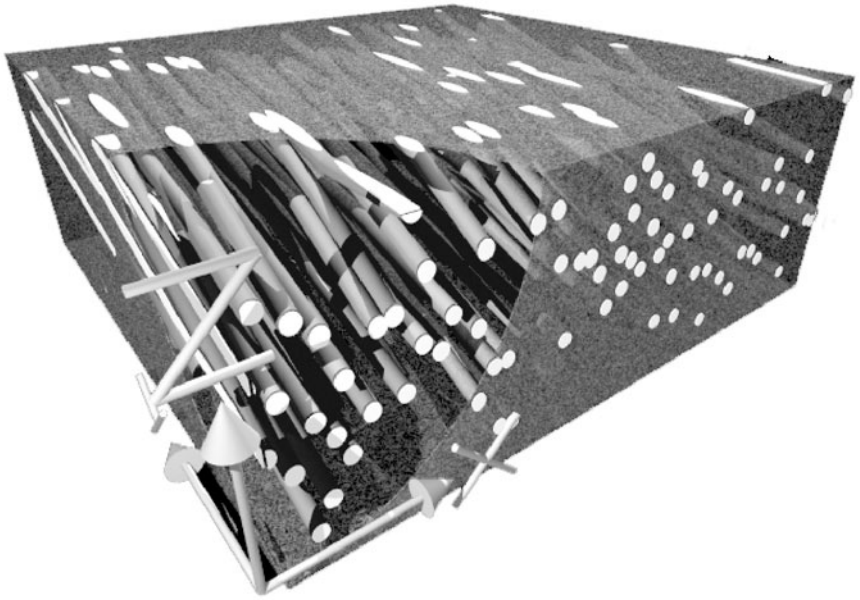


Figure 4.26 A volume rendering reconstruction of a glass fibre-reinforced polymer composite. The volume dimensions are 1.5 mm × 2.0mm × 0.5 mm in XYZ.

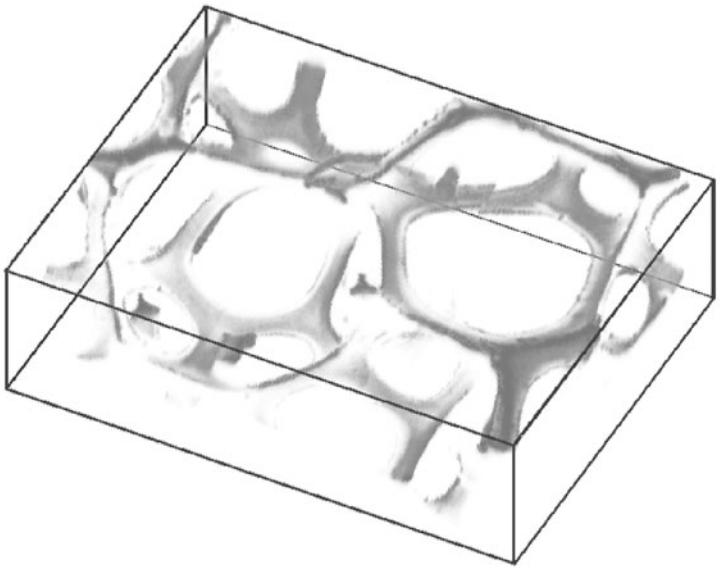


Figure 4.27 A 3D rendering of a polyurethane foam sample, which has been imaged in reflection mode. The volume dimensions are 2.5 mm × 1.5 mm × 1 mm in XYZ.

be amassed over time. Furthermore, much of the data will be redundant, if quantitative analysis techniques are applied later. For studies involving the analysis of multiple regions or large areas, alternative sampling techniques are more favourable.

4.5.3 Sectioning schemas

The optical sectioning of the CLSM suggests two possibilities for sampling and these are shown in Figs 4.28 and 4.29. The most appropriate method depends on the measurements required. If the features to be sectioned are approximately parallel to the microscope stage, for example when 3D fibre waviness within unidirectional fibre-reinforced composites is being studied (see Section 4.7), XZ frames taken at different Y are most appropriate. This sectioning scheme is shown in Fig. 4.29. Here the sample has been sectioned so that the fibres are lying essentially parallel to the sample surface. This method enables the expected low level of fibre waviness in Z to be followed, in individual fibres, over large distances in X and Y. However, if the features of interest are aligned with the optical axis of the microscope, it would be better to take XY frames at two or three Z positions and move over large areas in XY (as shown in Fig. 4.28).

4.5.4 Surface studies

As an object passes through the plane of focus, the reflected light intensity shows a sharp maximum. This effect can be seen clearly from the axial response curves plotted by measuring the image intensity as a mirror is passed through the

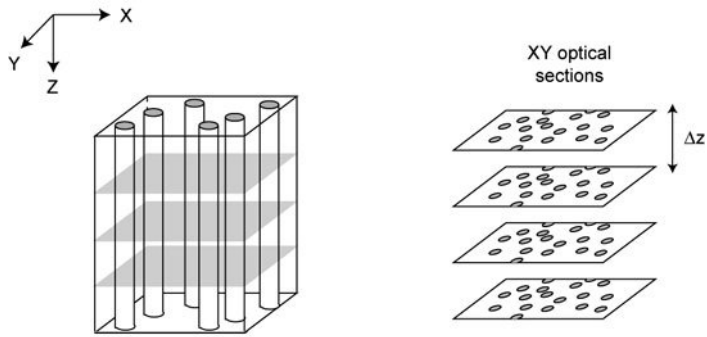


Figure 4.28 A schematic of one technique for 3D reconstruction using a CLSM system is shown. A series of XY optical sections are taken, each separated by Δz. By pattern matching object centre coordinates on one plane with those centres on the next plane, the orientation of fibres or the size distributions of objects may be deduced.

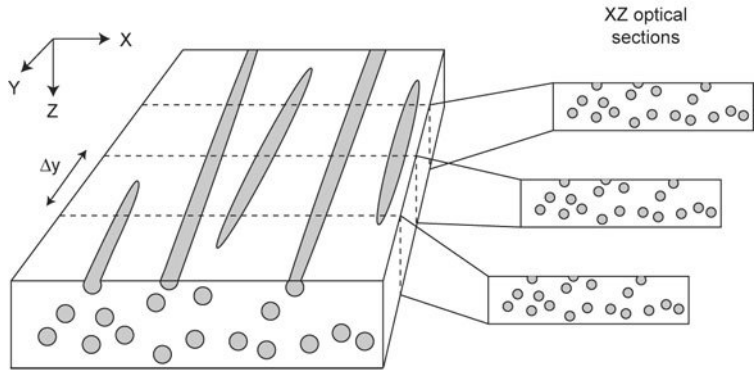


Figure 4.29 Another technique for 3D reconstruction using the CLSM involves taking a series of XZ sections, where each section is separated by Δy. If the sample is cut as shown, with the fibres lying almost parallel to the section plane, the fibre cross-sections will appear as near-circular images in the XZ planes. By pattern matching between XZ sections, the 3D waviness of fibres over significant fibre lengths may be determined.

focus, as shown in Fig. 4.6. For objects with more complex surface topologies, intensity maxima occur at different Z locations for different regions of the image, depending on the height of the sample at that particular point. A measure of the stage position at the occurrence of the intensity maximum enables the surface level to be recorded at each point across the field of view. In practice a number of XY images are acquired at regular increments in the Z direction. During the acquisition of this image series, a height image is created which retains the Z location at which the maximum intensity was found for each pixel. Figure 4.30 shows a height image of a silicon chip where the acquisition time for this image was approximately 20 seconds.

The precision in measurement of surface height depends on the microscope axial response and is therefore dependent on a number of factors (which were discussed in Section 4.3). The NA of the objective lens is one of the most significant factors. It was seen in Section 4.1.2 that the axial resolution of a microscope can be determined by the application of the Rayleigh criterion, i.e. the Airy patterns formed by two point objects, separated by some distance in the axial direction, must be separable. For a confocal microscope, fitted with an oil immersion objective lens, this resolution can be as low as 0.5 μm. However, the axial resolution for locating a surface (as opposed to resolving beads) depends on the accuracy with which the peak intensity of a single Airy pattern can be determined. This has been shown, experimentally, to be less than 0.1 μm.[28] In practice, the precision is often limited by the precision of the mechanical system providing the Z movement. The Noran Odyssey uses a stepper motor that drives the fine focus control of the microscope. The smallest step size with this device is 0.05 μm. Figure 4.31 demonstrates the accuracy and repeatability of this

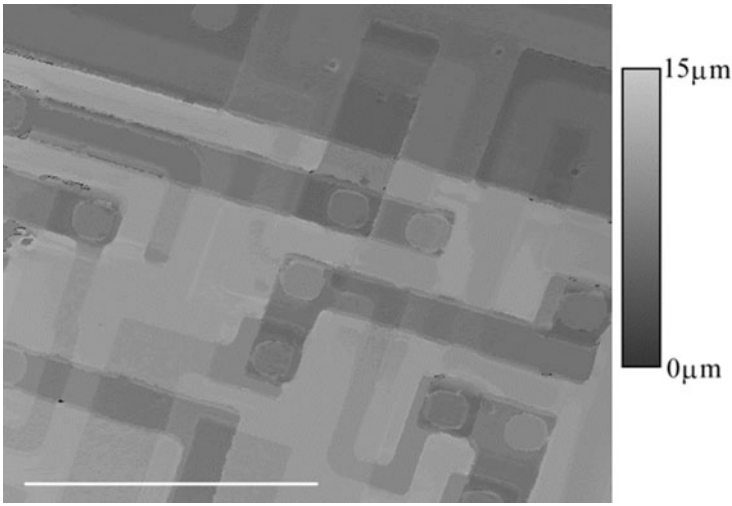


Figure 4.30 A depth map of a silicon chip. The image scale bar represents a length of 50 μm. The gradient scale bar shows the height represented by each grey level. Regions where it was impossible to identify the surface location are highlighted in white.

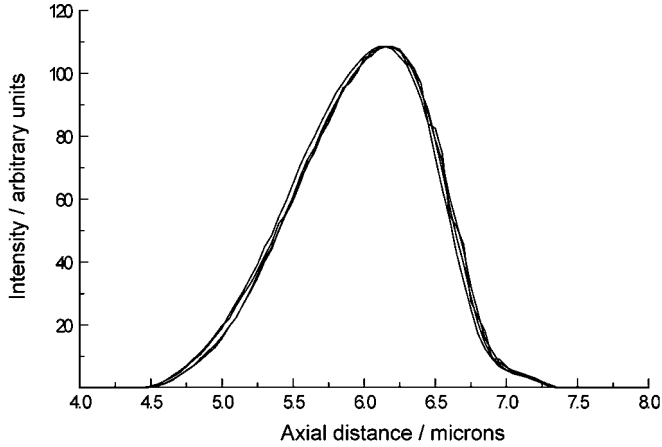


Figure 4.31 The repeatability of the z-drive is indicated by acquiring the same surface profile of a mirror slide a number of times.

system. The axial response function of a mirror has been plotted four times, with the stepper motor returning to the origin after each plot. The peak position is reproducible to an accuracy of $\pm 0.05\,\mu$m.

The intensity of the reflection at a surface is dependent on the refractive index difference between the surface and the medium in which it is immersed. If the refractive index difference is small, very little light is reflected at this

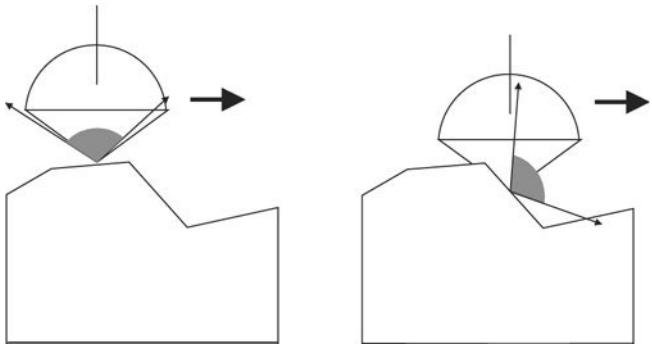


Figure 4.32 A schematic representation of an objective lens focussing along a surface with complex topology. It can be seen that, as the surface inclination increases, less of the light is reflected back into the objective lens.

interface. Either a higher gain setting must be selected, which increases noise, or the sample must be prepared in such a way as to increase the reflected light intensity. A suitable preparation used in a number of other surface microscopy techniques is to metallise the surface of the sample, effectively creating a mirrored surface.

The reflected light intensity is dependent on the tilt angle of structures on the sample surface, which can cause problems if samples with rough topologies are analysed.[29] When the tilt angle of the surface is low, a large percentage of the reflected light is within the solid angle of the objective lens. If the tilt angle is high, a much lower percentage of the reflected light is within the solid angle of the objective lens, as illustrated schematically in Fig. 4.32. For samples with very rough surfaces, the range of surface reflection intensities may make it inappropriate to use a single PMT gain setting. There is, of course, an ultimate limit to the steepest surface gradient, ϕ_{max}, from which a detectable signal is reflected, which is related to the vertex angle of the light cone as follows, $\phi_{max} = \alpha$. Therefore, objective lenses with higher NAs are more suited to analysing samples with rough surfaces. However, there will be slopes which are outside the collection range of the objective lens, i.e. $\phi_{max} > \alpha$. This effect is demonstrated in Fig. 4.33 where a cluster of small spherical beads have been imaged. There are regions within this image where the reflected light has not been collected and, therefore, where a height has not been assigned. These 'holes' can be artificially filled by averaging neighbouring pixel intensities,[30] although this method is clearly undesirable. Figure 4.34(a) shows the surface of a thin film that has been located using the Noran Odyssey. The film was metallised by sputter-coating with aluminium in order to increase the intensity of the reflected light from the surface. Figure 4.34(b) shows a 3D visualisation of the same surface (using the surface rendering technique described in Section 4.5.2), with specular, ambient and diffuse lighting models.

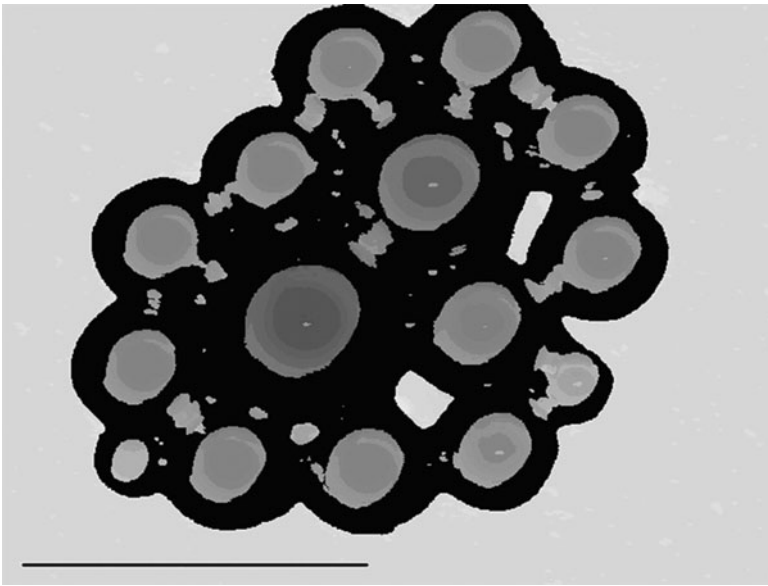


Figure 4.33 A reflection mode surface image of a group of metallised latex spheres. The scale bar represents a length of 20 μm.

4.5.5 Auto-focus imaging

With a conventional microscope, an image of a surface with complex topology will consist of regions that are in focus and regions that are out of focus. With a confocal microscope, the regions that are out of focus will not be visible within the image. In some applications, it may be desirable to image the sample in its entirety while maintaining good focus throughout. In order to achieve this aim with either a confocal or a conventional microscope, a series of images must be collected with a range of z locations. The method for combining the images in this series to form an auto-focus image depends on the microscopy technique used.

In conventional microscopy, each image must be processed in turn in order to identify the regions that are in focus. These in-focus regions are then collected and combined to form the auto-focus image. The image processing requirements of this technique are non-trivial; however, there are commercial packages which perform these tasks very well,[31] although the speed of operation is quite poor.

In confocal microscopy, the out-of-focus regions from each image have already been removed by the microscope optics. Therefore, the only processing required is the combination of the image sections. This can be performed on-the-fly (i.e. in real time) as the image sections are being captured. An auto-focus image is created by recording the maximum intensity at each pixel location, as

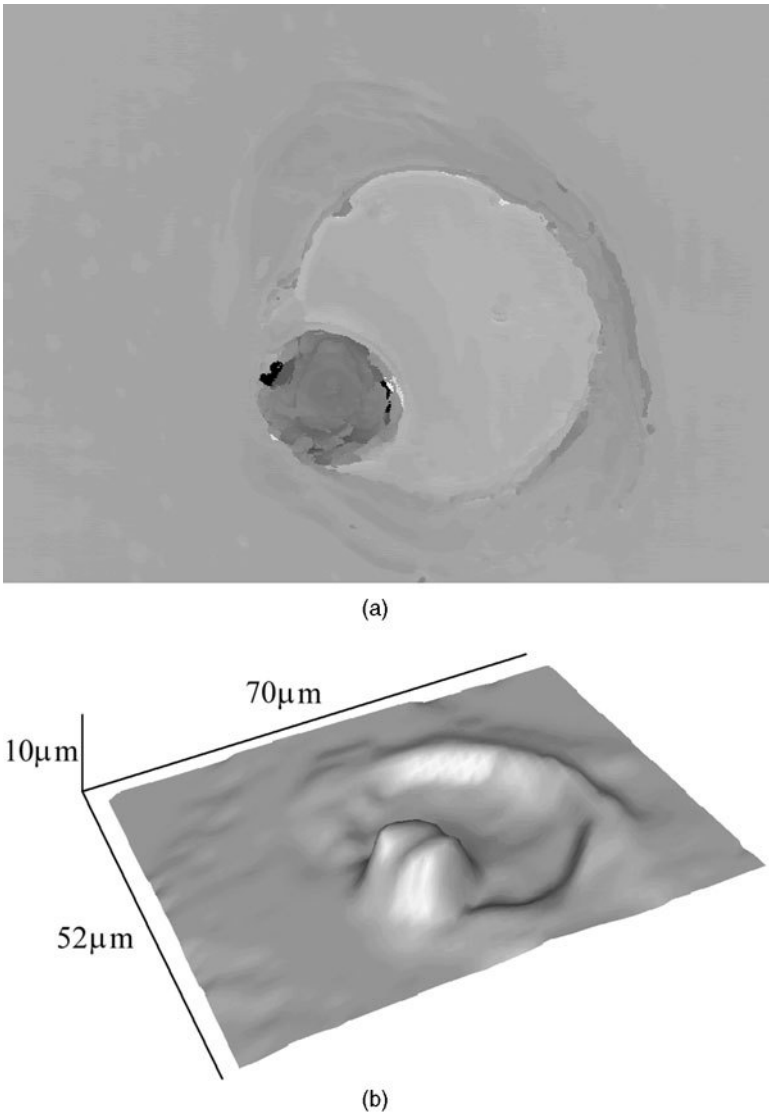


Figure 4.34 The surface profile of a thin polymer film containing small glass inclusions was required. During the draw process, the bead has broken the surface of the film. (a) A surface profile image is shown and (b) the 3D rendered surface for comparison.

the surface is sectioned. The auto-focus routine is almost identical to the surface topology routine, but the intensity is recorded rather than the z location. Figure 4.35 shows, as an example, an auto-focus image of a polyurethane foam sample which was acquired over a range of $\Delta z = 300\mu$m.

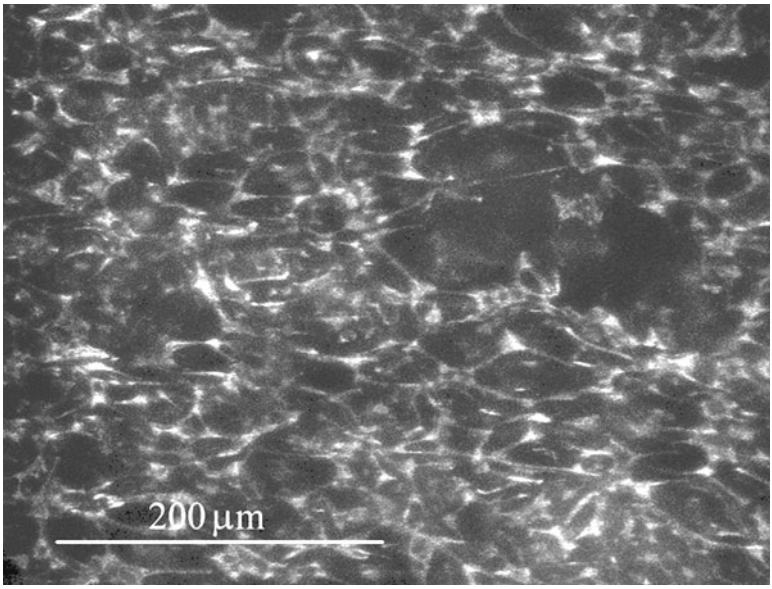


Figure 4.35 An auto-focus image of a polyurethane foam sample, taken using Nikon PlanApo ×10 air objective lens. This image was acquired over a range, $\Delta z = 300\,\mu$m.

4.6 Case study: thin film particulate analysis

4.6.1 Summary

This case study describes the measurement of particulate number and volume statistics within 3D volume datasets. The problems of accurate volume measurements and unbiased counting using confocal microscopy are addressed. To illustrate the technique, two thin polymer films containing fluorescent-labelled, anti-blocking silicas are analysed and the differences between their particulate statistics are highlighted.

4.6.2 Introduction to polymer films

By definition, a polymer is a large molecule (macromolecule) composed of a repeated sequence of smaller chemical units. These long molecular chains may be linear, branched or cross-linked and may be up to 10 μm in length. In the molten state, the chains are flexible and adopt certain orientations depending on external forces applied to the bulk material (such as shearing and tension). The organisation of the macromolecules has a significant effect on the physical (including optical) properties of the bulk material.[32]

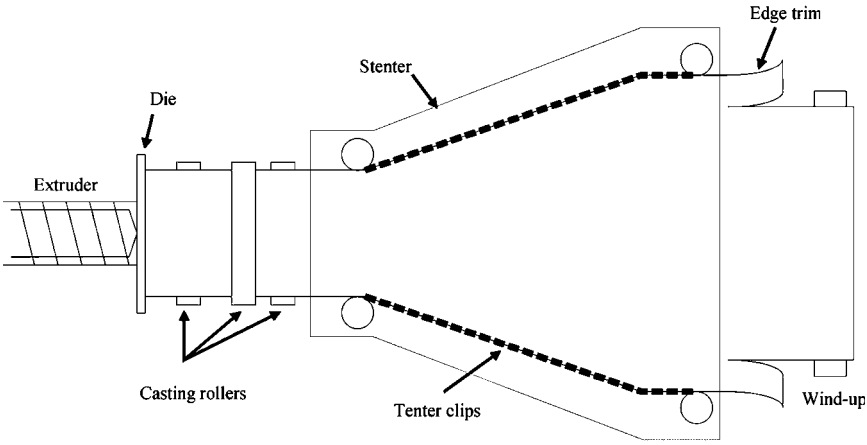


Figure 4.36 A schematic illustration of the industrial production of polymer film by extrusion and bi-axial drawing.

Polymer films have a number of useful characteristics, such as high tensile strength, high resistance to tearing and abrasion, good electrical insulation, and excellent optical properties. These qualities have ensured their success in a large number of industrial and domestic applications. To illustrate with a single application, consider a polymer film that has been used as the base for magnetic recording cassette tapes. The film's intrinsic strength means that thinner films are possible, allowing for greater storage within the same volume of tape. The good creep recovery of the film means that it will not lengthen with age. Finally, the film is very cheap to manufacture.

The industrial manufacture of a bi-axially drawn polymer film is illustrated schematically in Fig. 4.36. Polymer chips are fed into the extruder where they are heated, then extruded onto a cooled casting drum. The film is heated and stretched in the longitudinal direction to approximately three times its original length by a series of rollers running at different speeds. Following this, the film is held by two sets of continuous clips, which grip each edge. These clips carry the film into the stenter, which heats the film and stretches it in the transverse direction to approximately three times its original width. While under tension the film is crystallised in a series of specially designed ovens, after which the film is finally cooled and wound into rolls.

During packaging and the high speed rolling process the film is continually in contact with its neighbouring sheets. The smooth film surface results in a large area of contact which causes them to stick together (sometimes referred to as blocking). In order to reduce the area of contact, small particulates (anti-blocking agents) can be added to the polymer film during the manufacture of the raw polymer chips. The refractive indices of these additives are chosen to closely match that of the polymer so as not to affect the film clarity. As the film is drawn, the hard particulates will cause the formation of small asperities (i.e.

bumps) in the film surface. The resulting increase in surface roughness will improve the film handling properties.[33]

4.6.3 Stereological analysis of particulates

The analysis of discrete particulates is a common problem in microscopy and whether the particles are cells, grains of sand or, in this case, silicas, similar stereological tools can be applied to all. This case study requires the measurement of two distinct properties: particulate density and volume distributions.

Before the advent of 3D microscopy, particle number measurements were usually performed by sectioning the sample. An unbiased particle counting procedure, the disector, is illustrated in Fig. 4.37 where a pair of registered serial sections, separated by a known distance *h* are randomly selected. The counting frame described in Section 2.9 is used to identify particles in the first section, termed the reference section. An estimate of particle number density can be made from a count of objects that are present in the reference section but are absent in the look-up section. This procedure is repeated over a number of randomly located serial sections until sufficient statistical accuracy has been achieved. It is clear that the manual disector method is time-consuming, although recent developments in image processing may enable the automation of this technique. Furthermore, the development of a novel tandem projection microscope, which enables simultaneous viewing of serial sections, can give a significant increase in speed.[34]

The measurement of particulate volume from a number of serial sections can be performed by application of the Cavalieri technique.[35] The area of the exposed surface of the particulate is measured on each section. The best estimate

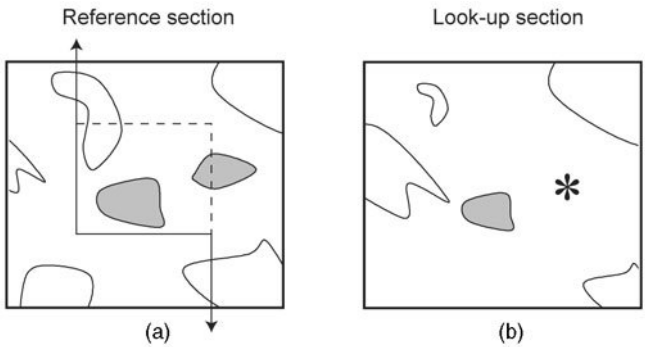


Figure 4.37 The disector principle: (a) objects are only counted if they appear within the unbiased counting frame in the reference section but do not appear in the look-up section illustrated in (b), which is parallel to the reference section plane. Only one of the two shaded objects would be counted.

volume is determined by summing the area of each intersect multiplied by the section spacing. The Cavalieri, like the disector technique, is also very time-consuming to implement.

In summary, it is possible to derive an unbiased particle size distribution and particle number from a set of serial sections of known separation by a combination of the disector and Cavalieri techniques. Although these techniques give good results, there are many situations or object configurations that will give rise to errors. Some examples of 'problem' cases are objects that are closer than the section spacing, objects that are smaller than the section spacing and concave objects. A possible solution to the first two problems is to decrease the section spacing distance. However, in practice, constraints imposed by the sectioning equipment, and the time taken to analyse the increasing number of sections, imposes strict limitations on this technique. Recently, researchers have been favouring the confocal technique because of its optical sectioning capabilities.

4.6.4 Confocal microscopy of particulates

Both the disector and Cavalieri techniques require the acquisition of a number of perfectly registered thin sections. The confocal microscope is an ideal tool for such measurements. The disector principle is based on the collection of discrete sections, whereas the confocal microscope allows a continuous sectioning throughout the depth of the sample. To take advantage of this feature, a new sampling technique called the 'optical disector' was proposed, which leads to a significant increase in practical efficiency. The optical disector uses a 3D unbiased counting frame termed the 'unbiased brick', illustrated in Fig. 4.38. With this 3D counting frame, there are three acceptance surfaces and five

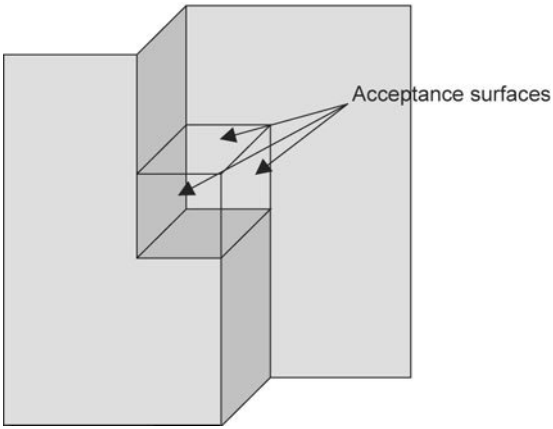


Figure 4.38 The 3D analogue of the unbiased counting frame used in 2D images is called the unbiased brick.

forbidden surfaces and similar rules as for the 2D case are applied. The microscopist moves through the sample and manually counts objects using a strict set of rules.[36]

With modern commercial confocal microscopes, the addition of a host computer enables the collection and storage of 3D volumes, which is now a standard feature with most models of CLSM. Three-dimensional image processing of reconstructed volumes enables an accurate estimate of particle number, without any assumption of object morphology. This is in contrast to the disector and optical disector methods which both assume convex objects. An example of this method is given by Karlsson and Liljeborg,[37] who used this reconstruction technique to measure the arrangement of pores in translucent alumina.

The volume of data that may be rapidly acquired by the confocal technique removes the need for many of the stereological counting tools, but there are some drawbacks to using the confocal technique. The appearance of a particular section is now dependent on the illumination path, which has a significant effect on the image (as described in Section 4.3.4). Also, inhomogeneous sample fluorescence and reflectivity may further degrade the image quality.

The subject of light attenuation correction in confocal microscopy has only recently been tackled. Liljeborg *et al.* describe a 'histostack' method for light attenuation correction where the attenuation factor of the sample is calculated from section histograms.[38] The attenuation is then compensated in each XY section. Margadant *et al.* describe a computationally intensive attenuation correction which calculates the attenuation at each individual image voxel (the abbreviated term for volume element).[39] A technique is described here which compensates for light attenuation and more importantly (in this particular application) inhomogeneous sample fluorescence.

Figure 4.39 illustrates the problem of inhomogeneous sample fluorescence. High PMT gain and low PMT gain images from a sample of low density, polyethylene (LDPE) film containing fluorescent silica particles are shown. Inspection of the low gain image, Fig. 4.39(a), shows that only the largest of the silica particles are visible, whereas in the high gain image, shown in Fig. 4.39(b), a large number of smaller particles can be observed. Inspection of the larger particulates in the high gain image shows that there is considerable blurring of these objects along the optical axis. Therefore, particulate size and number measurements will be under-estimated if low gain images are considered and over-estimated for high gain images. Note that this problem cannot be solved by Liljeborg's 'histostack' or Margadant's voxel approach.

4.6.5 Volume measurement methods

During manual operations of a microscope, it is the job of the microscopist to select a suitable gain for a particular image. This gain setting will usually be

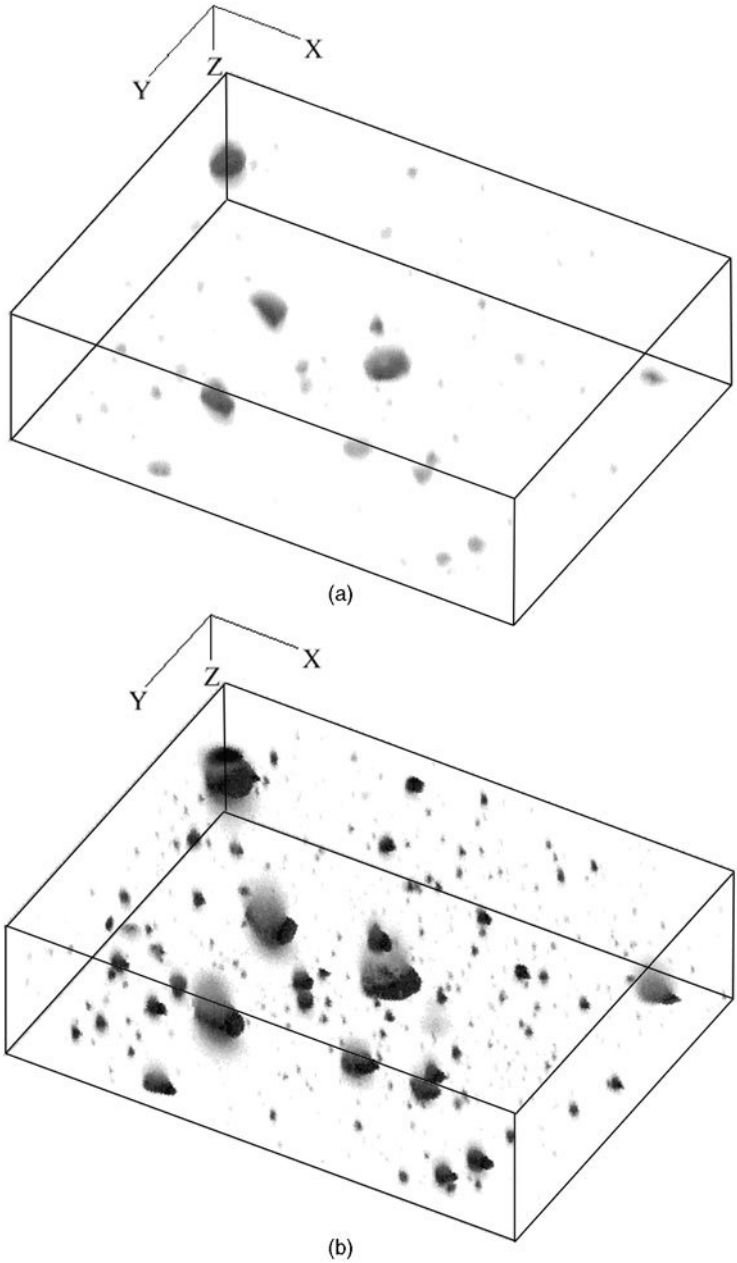


Figure 4.39 Two reconstructions of fluorescent particulates within a polymer film are shown. The larger particulates are visible in the lower intensity reconstruction in (a) but the smaller particulates cannot be seen. However, the higher gain reconstruction in (b), makes visible the smaller particulates, but at the expense of blurring the larger particulates.

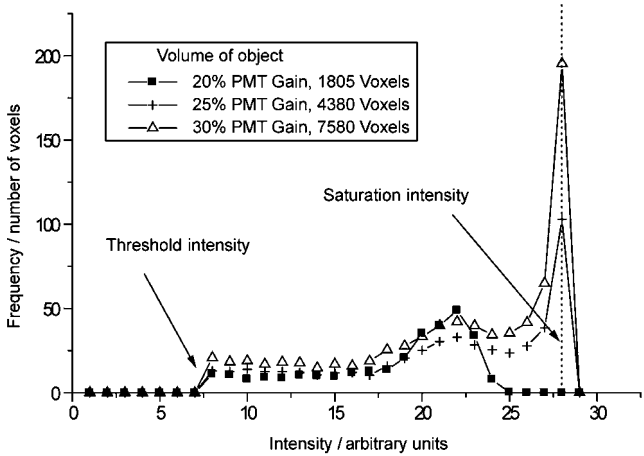


Figure 4.40 The intensity histogram and volume (in voxels) of a single object at different gain settings.

determined by inspection of the objects within the field of view. A suitable gain will give objects with 'healthy' intensity histograms, i.e. an intensity range above the background level but with minimal saturation, which closely matches the dynamic range of intensities present within the field of view. It is clear from Fig. 4.39 that, in some cases, a gain setting cannot be found which gives a 'healthy' histogram for all the objects within the volume. Therefore, a reconstruction of such a volume requires the collection of a number of datasets at different gain levels, together with a recombination algorithm.

Six datasets of the region illustrated in Fig. 4.39 were collected at regular gain increments. The datasets were segmented by application of a single threshold and the objects within these volumes were identified using the voxel connectivity, the 3D equivalent of the pixel connectivity algorithm described in Section 2.8.6. The centroid and volume of each object, together with their intensity histograms (in 3D) were recorded. Figure 4.40 shows the intensity histogram for an object viewed at three different gain settings. As the gain is increased, there is an apparent increase in the size of this object. There is also an increase in the number of saturated voxels. By inspection of these histograms, it would appear that a gain of 20–25% is most suitable for this particular object.

An algorithm was written that enables the automatic selection and recombination of objects from any number of datasets acquired at exactly the same position within the sample but with different gain levels. The algorithm considers each object in turn and decides whether the object should be placed into a new 'composite' dataset. The decision process is illustrated as a flow diagram in Fig. 4.41. The main body of the algorithm is based on the voxel connectivity routine described earlier. When a new object has been identified, a set of rules is applied to determine whether it should be added to the composite

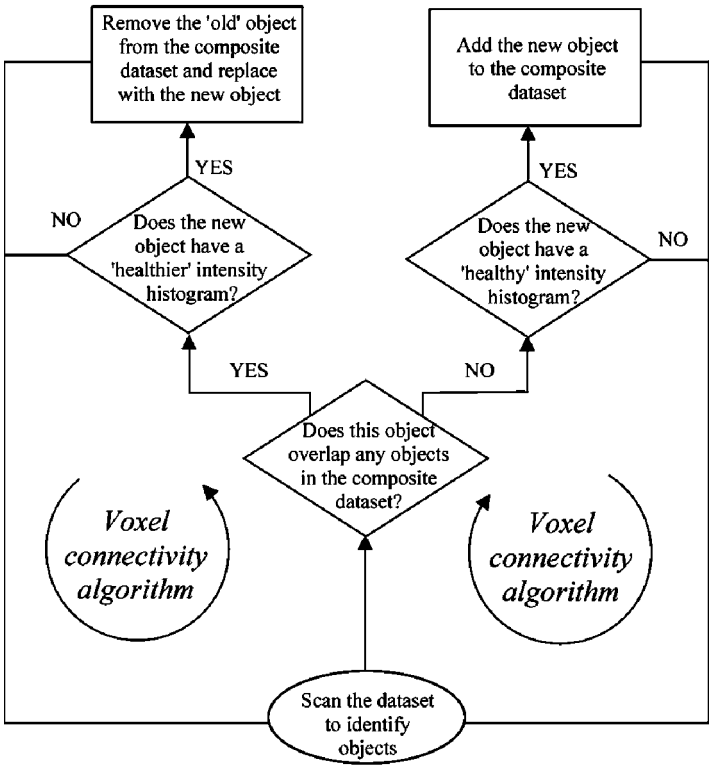

Remove the 'old' object from the composite dataset and replace with the new object
Add the new object to the composite dataset
YES
YES
NO
Does the new object have a 'healthier' intensity histogram?
Does the new object have a 'healthy' intensity histogram?
NO
YES
NO
Does this object overlap any objects in the composite dataset?
Voxel connectivity algorithm
Voxel connectivity algorithm
Scan the dataset to identify objects

Figure 4.41 A flow chart illustrating an image processing scheme that can be used to combine objects from volume datasets.

dataset. Objects are added or removed from the composite dataset by a recursive 'flood fill' algorithm as described in Section 2.8.6.

The algorithm relies on an evaluation of each object's histogram to determine whether it should be placed into the composite dataset and, in the cases of overlap, whether it should replace an object already in the composite dataset. A simple evaluation scheme that compares the mean voxel intensity and the percentage of saturated voxels in each object was adopted. Figure 4.42 shows a 'composite' volume, which is a recombination of six datasets acquired at different gain settings. It is clear from Fig. 4.42 that both large and small objects are present within the 'composite' dataset. Furthermore, there is no noticeable blurring of the larger objects, and sections through the objects, reveal only small amounts of intensity saturation.

Segmented images formed by the application of a single intensity threshold often exhibit clustering of smaller particulates around the larger bodies. These 'satellite' objects are usually the result of image noise. An illustration of how these artefacts develop in 2D images is shown in Figs 4.43(a) and (b), where an irregular object boundary has occurred due to the spatial noise present in the

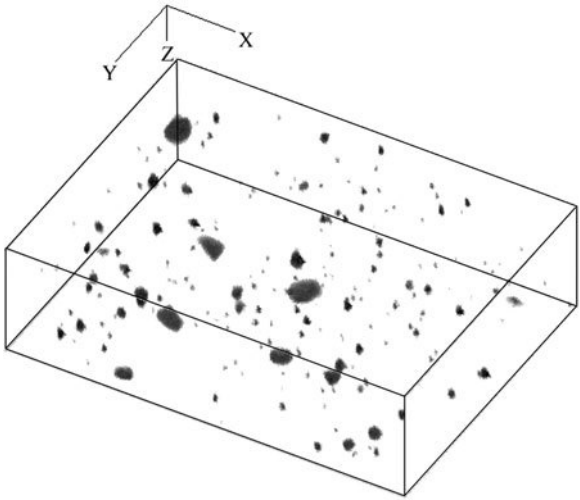


Figure 4.42 A composite volume reconstruction has been created from six datasets which were acquired at different gain settings.

image. Close inspection of the binary image (b) shows that, at the boundary, there are some pixels which are not joined to the main body of the object. If a pixel connectivity routine were applied to this image, these 'fragments' would be counted as separate features, resulting in satellite objects around the larger feature. Some researchers (for example, Liljeborg *et al.*[38]) use spatial filtering to reduce the noise in the image. Figures 4.43(c) and (d) show the same object after smoothing by a 3×3 mean filter. Although this technique produces smoother objects and reduces the probability of satellite objects, spatial filtering should be avoided if possible because it ultimately leads to a reduction of resolution and image contrast. A more desirable approach is to eliminate the 'satellite' objects by inspection of their intensity histograms. These objects have low intensity

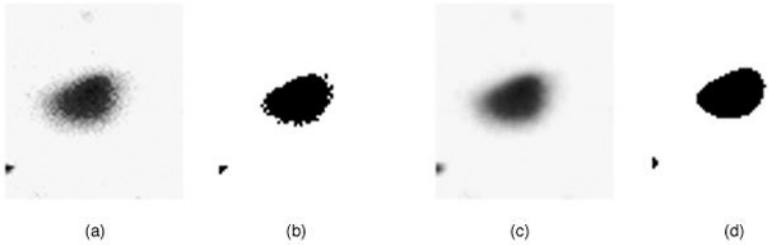


Figure 4.43 The noise surrounding a small object imaged in (a) can result in the presence of satellite objects after segmentation, as illustrated in (b). Smoothing of the original image with a 3×3 neighbourhood averager filter, as shown in (c) is one method which can be effective in removing these satellite objects, as in (d).

maxima, which are usually only slightly higher than the threshold intensity and, hence, their intensity maxima can be used to filter out the satellite objects from the 'composite' dataset.

4.6.6 Experimental

The distribution of anti-blocking silica particulates within two different LDPE film samples, named sample A and sample B, were analysed. The silicas were labelled with Rhodamine 6G, a commonly used dye, which has an absorption peak at 530 nm and emits a broad band of fluorescence at higher wavelengths. Small pieces of each film sample (approximately 2 cm $\times$ 2 cm) were mounted in an immersion oil with a refractive index of 1.5, providing a close match to the refractive index of the LDPE. Although the good refractive index match minimises aberrations (see Section 4.3.2), it does however mean that there is little or no light reflected at the oil–film interface. In order to make particle density calculations, it is essential that the reference volume is measured and this necessitates the location of both the top and bottom film surfaces. The top and bottom surfaces of the film samples were lightly sputter-coated with aluminium so that they were visible in reflection mode. This light coating did not have a significant effect on the sub-surface fluorescence intensity.

The variable gain technique described earlier was used to acquire twenty volume datasets at random locations within each sample, one of which is illustrated in Fig. 4.42. Each volume was also imaged in reflection mode in order to identify the top and bottom surfaces. The depth of these datasets was greater than the film thickness for both samples to ensure that the entire film thickness was always present. A projected version of the unbiased counting frame, with width *w* and height *h* as illustrated in Fig. 4.44, was applied. This method is slightly simpler than the unbiased brick counting frame (see Fig. 4.38) that is usually employed in 3D counting applications.

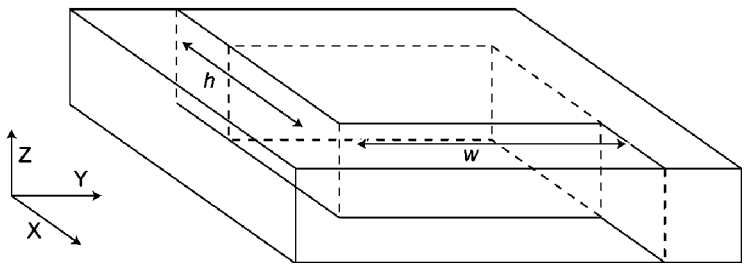


Figure 4.44 A 3D unbiased counting frame, width *w* and length *h*, formed by two parallel planes of separation *t*.

4.6.7 Results and conclusions

The mean film thickness was measured from the volume dataset of each film and the thicknesses were $t_A = 41\mu$m for sample A and $t_B = 23\,\mu$m for sample B. The particulate number count, n, can be converted to particulate density, N, as follows,

$$N = \frac{n}{w.h.t} \qquad [4.11]$$

where t is the film thickness, and h and w are the dimensions of the counting frame. The particle density distribution for each sample is shown in Fig. 4.45. It is clear from this plot that sample B has a higher particle density with a silica volume fraction of 0.0023% compared to sample A, which has a silica volume fraction of 0.0011%. However, it is not clear from this figure whether the samples exhibit different particulate volume distributions. A clearer comparison of the two particulate volume distributions can be seen in Fig. 4.46 where the two distributions have been normalised. This plot shows the relative number of particulates found within particulate volume ranges (or 'bins'). Hence, it is termed a number-weighted distribution. However, it is difficult to compare the two distributions in the range $> 0.5\mu$m. An alternative method of plotting the same data is shown in Fig. 4.47. Here, each column has been weighted by the mean volume of the 'bin', giving a volume-weighted distribution. Both number-weighted and volume-weighted plots are equally valid methods of displaying the data, however the method of plotting *must* be recorded. From the volume-weighted graph, it can be seen that the particulate volume distribution is quite similar for each of the two samples. It is certainly possible that the anti-blocking silicas in each sample came from the same source but with different concentrations.

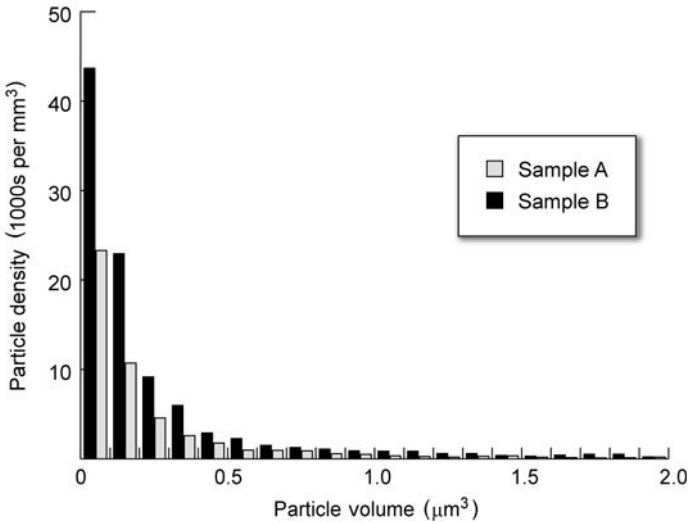


Figure 4.45 The particulate density distribution for two different samples.

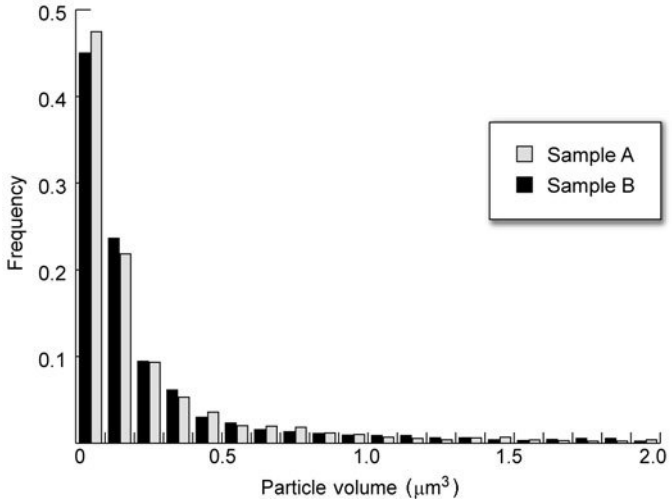


Figure 4.46 The normalised, number weighted particulate frequency distribution.

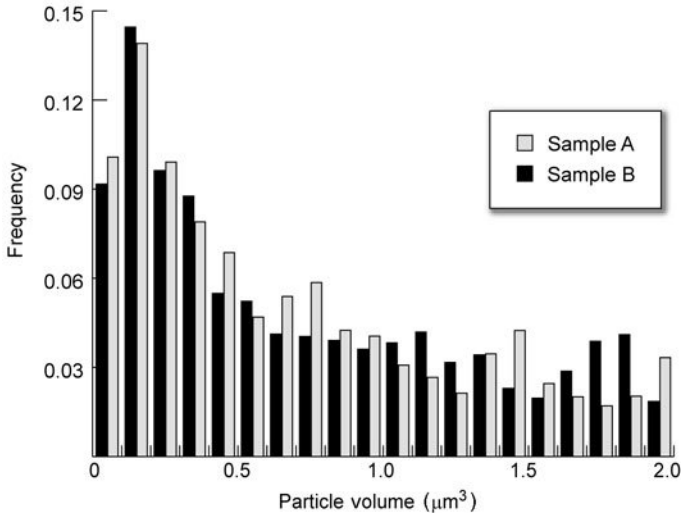


Figure 4.47 The normalised, volume weighted particulate frequency distribution.

4.7 Case study: fibre waviness

4.7.1 Summary

The first direct, unambiguous measurement of individual fibre waviness has been achieved within glass fibre-reinforced, polymer composites using confocal laser scanning microscopy. A novel technique is outlined for finding and fitting

the best curves to the 3D fibre data, which also has the merit of smoothing out intrinsic, high frequency instrumental noise. Hence, the determination of local curvature and torsion of the fibres within composites or textiles or other fibrous materials becomes possible.

4.7.2 Introduction

The physical properties of a material determine how it reacts and deforms when under the influence of an external force or forces. If these properties are determined accurately, it is possible to predict how a component made from this material will perform in a specific application. It is this knowledge of physical properties that enables engineers to choose the correct material for a particular application. In general, most mechanical applications involve elastic deformation of the component parts. For example, the frame of a tennis racket will deform slightly as the ball is struck and, shortly after the initial deformation, the frame returns to its original shape, i.e. the deformation was elastic. However, when a material is loaded excessively, it fails (breaks). In all mechanical applications, engineers must consider the implications of the materials failure strength (especially in critical applications, such as bridges).

The elastic behaviour of most materials is predictable, for example a spring will react in exactly the same way each time it experiences the same compression. This is because during low-strain elastic deformation, the bonds between molecules are stretched but do not break, hence the structure of the material does not change. The physical strength of a material is somewhat harder to determine and accurately predicting the point of total failure is not always possible. The onset of failure occurs when the strained material starts to fracture and the rate of fracturing increases until total failure occurs. The statistical nature of failure is such that it is not possible to predict exactly when a fracture will occur. The strength of a material is always that of its weakest part, therefore fractures tend to initiate around defects.

The maximum compressive strength of a unidirectional fibre-reinforced composite is parallel to the fibre direction. When slight fibre misalignments are present within a sample, the stress tends to build up in these areas. This stress further misaligns the fibre and its neighbours resulting in the formation of a kink band, as illustrated in Fig. 4.48. As the kink band grows the fibres fracture and the entire load is borne by the matrix, resulting in total failure, as seen in Fig. 4.49. It is believed that fibre waviness is the primary cause of failure in compression. A number of different models, which relate fibre misalignment to failure stress, have been developed. Although they differ in specifics, many contain as one of the terms, the reciprocal of fibre misalignment. For example, Argon proposed that the compressive failure stress, σ_c, is related to the matrix shear yield stress τ_{Ym} and the fibre misalignment angle $\Delta\phi$ as follows:[40]

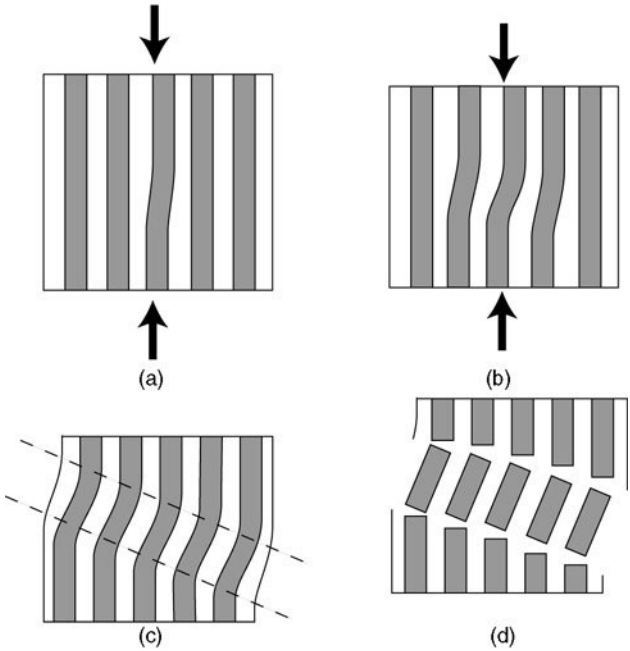


Figure 4.48 (a) The unidirectional loading of a unidirectional composite sample with a single misaligned fibre is shown. Initially, the loading results in kink band formation, see (b) and (c), where neighbouring fibres are forced into misalignment. This effect will eventually result in the failure of the sample, as illustrated in (d).

$$\sigma_c = \frac{\tau_{Ym}}{\Delta\phi} \qquad [4.12]$$

where $\Delta\phi$ is in radians. More recently it has been shown both experimentally[41] and by modelling[42] that fibre misalignments of less that one degree have a significant effect on the onset of compressive failure.

Unidirectional composites are normally fabricated by manual 'lay-up' of thin ribbons (prepreg) containing continuous fibres surrounded by resin. When the lay-up is complete, the composite is heated in order to cure the resin. During lay-up, ply misalignments may occur resulting in large regions of fibre misalignment in the finished composite. Furthermore, plies may contain resin rich areas, which distort neighbouring plies, resulting in fibre waviness in the finished composite. The characterisation of fibre misalignment and waviness is vital in order to understand and predict the compressive behaviour of fibre-reinforced composites. Unfortunately, the measurement of small-angle fibre misalignments and waviness is not an easy task. A number of non-destructive and destructive experimental techniques have been proposed for measurement of fibre waviness, although all of them have had limited success when presented with complex components. Many of these techniques have only been able to

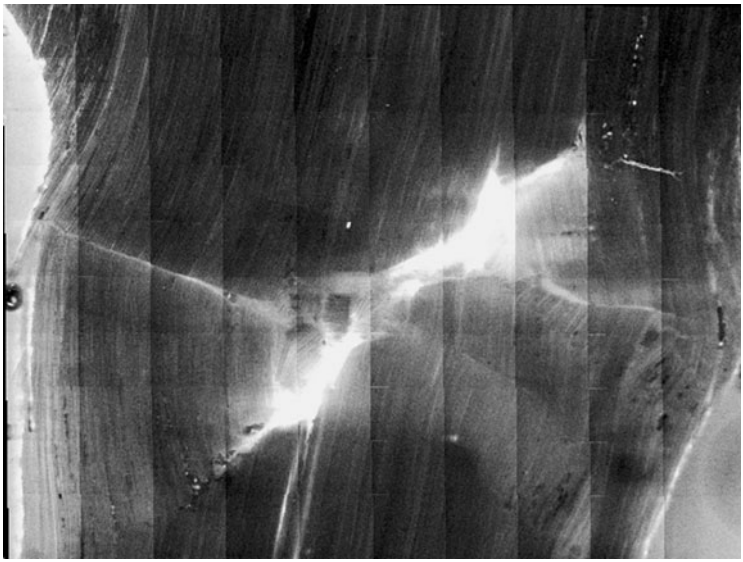


Figure 4.49 A large area, fluorescence mode image reconstructed from 100 XY overlapping image frames showing the compressive failure of a fibre-reinforced polymer composite.

give quantitative data on fibre misalignments for samples where the waviness has been artificially introduced during the fabrication process.[43] This artificially introduced waviness is high amplitude and coherent, i.e. many fibres are seen to exhibit cooperative movement. The most successful methods for measuring fibre misalignments and waviness are reviewed below.

A technique for the measurement of fibre orientations from a polished cross-section was described in a previous case study (see Chapter 3) and could be applied to identify fibre misalignments in unidirectional composites. However, this technique relies on the fibres having a perfectly circular cross-section and the accurate measurement of the cross-section ellipticity. The accuracy of orientation measurement can be improved by sectioning at an angle to the mean fibre direction, resulting in elongated ellipses.[44] However, it is questionable whether this technique can give sub-degree measurement accuracy.[45]

Fibre waviness can be observed in polished cross-sections parallel to the mean fibre direction (see Fig. 4.50), and misalignments measured directly from micrographs. Several techniques for coarse 'mapping' of fibre waviness have been based on this method.[46, 47] However, fibre waviness is not constrained to a plane and therefore more complex serial sectioning methods must be employed in order to fully characterise natural (process-induced) fibre waviness.[48] Clearly, serial sectioning is a complex and time-consuming task, with each section taking at least a day to prepare. Therefore, it is unlikely that anyone would perform such a characterisation on a routine basis. This case

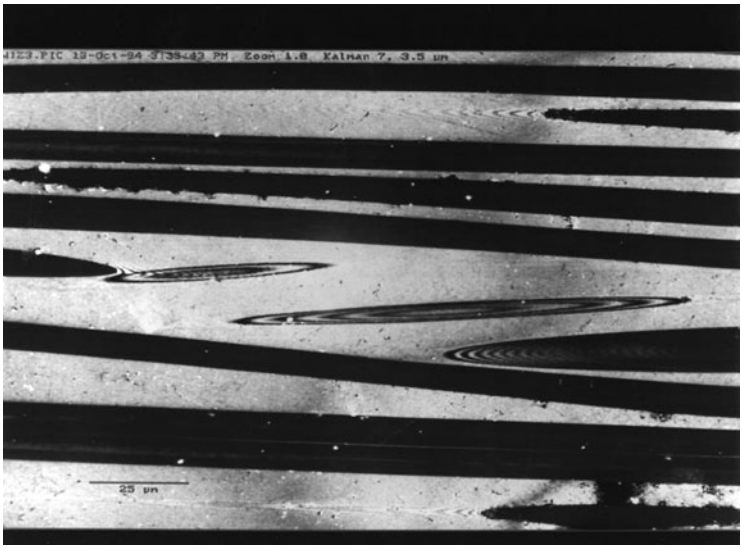


Figure 4.50 The XY dimensions of the frame are 200 μm $\times$ 150 μm. This image was taken at the surface of a unidirectional, glass fibre-reinforced composite. Most of the fibres lie in the section plane, but some are oriented at a small angle to the plane. Note that interference fringes due to light reflected off the surface interfering with the light reflected off the glass fibre as it disappears below the surface. The 'interference fringe islands' in the centre of the frame can only be due to two curved fibres just failing to break through the surface. Unidirectional fibres therefore exhibit local curvature and hence 3D waviness.

study describes a non-destructive sectioning technique using confocal laser scanning microscopy.

4.7.3 CLSM of unidirectional composites

The main drawbacks of manual serial-sectioning techniques for reconstructing fibre waviness are that it is very time-consuming and it is hard to achieve registration between successive sections. Confocal microscopy enables optical sectioning of the sample, which ensures the rapid collection of perfectly registered sections. Figure 4.51 shows a series of five (XY) images, acquired in fluorescence mode. Each image is separated by 10 μm. Therefore, the displacement of the fibre centres, from one image to the next, can be measured directly and, hence, their orientations can be inferred.

Figure 4.52 shows a pair of XZ images of a unidirectional composite (where the fibres are oriented in the XY plane). The sample surface is clearly visible in the reflection mode image, however the location of fibres beneath the surface is not an easy task. This is because the laser light, reflected at the fibre-matrix interface, will only re-enter the objective lens if it is within the collection cone

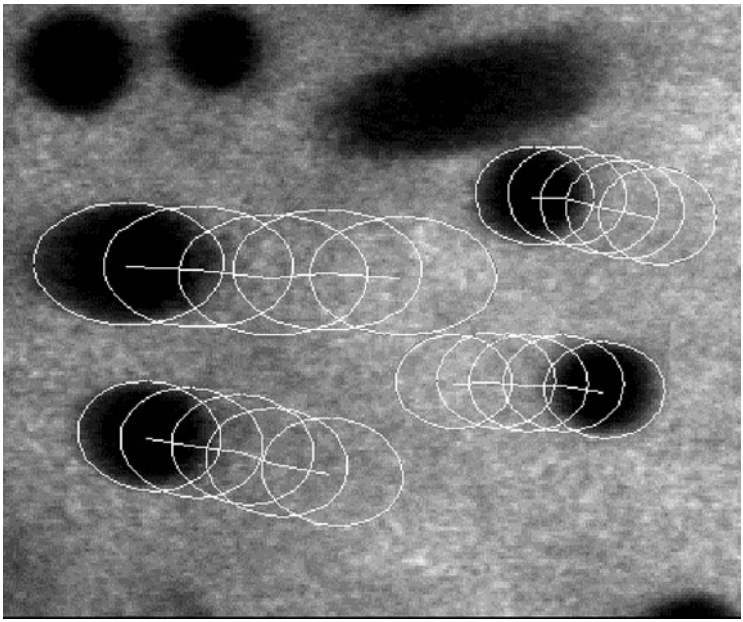
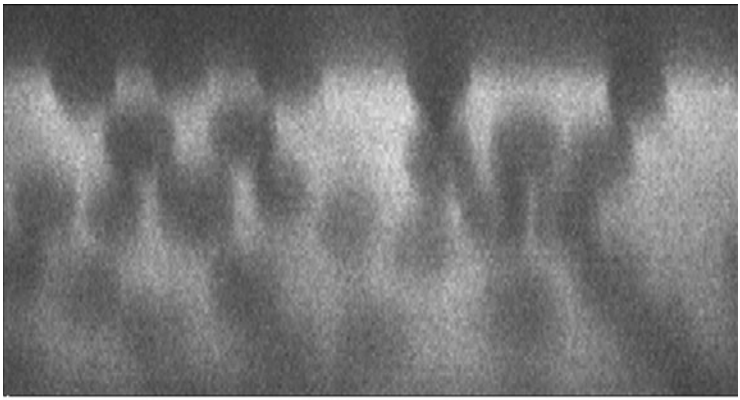


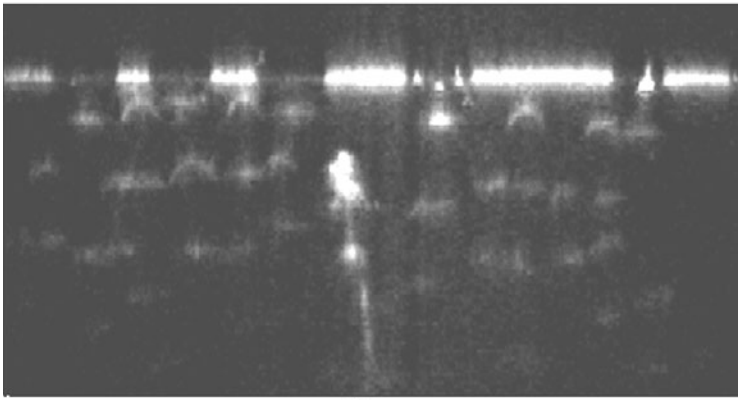
Figure 4.51 By taking a series of XY images at different Z depths and fitting the elliptical cross-section to each fibre's image, the shift in fibre centre coordinates can be measured. Hence, an independent estimate of the fibre orientation may be obtained. Note the fibre centre with the higher cross-section ellipticity has moved furthest, as would be expected.

(as described in Section 4.5.4). Therefore, only the tops and bottoms of the fibre appear on this image. When a sample fluoresces, light is emitted in all directions and hence the orientation of the fibre-matrix interface does not affect the image. However, in order to image a sample in fluorescence mode, it must fluoresce and fortunately most polymeric matrix materials have this property. Therefore, fluorescence is the most suitable imaging mode for identifying sub-surface features within fibre-reinforced composites.

The image quality of composite samples is highly sample dependent. Many factors affect image quality, although fibre packing fraction, extent of voiding and level of fluorescence of the matrix are the most important factors. When samples are sectioned optically, it is observed that there is a reduction in image quality with increasing depth, resulting in noisy images with poor contrast. This is partly due to the absorption and scattering of the laser light within the matrix, and more critically the fibre packing fraction, with the reflections at the fibre-matrix interface resulting in shadowing of features beneath. In practice it has been found that for high packing fraction samples, $V_f > 40\%$, optical sections taken at depths of $z > 40\mu m$ are so distorted and noisy as to hold no useful information. However, for low packing fraction samples, $V_f < 10\%$, images of



(a)



(b)

Figure 4.52 XZ sections of a glass fibre-reinforced polymer composite (a) in fluorescence mode and (b) in reflection mode. Both sections are 100 μm $\times$ 50 μm in XZ.

good quality may be acquired over the full working distance of the objective lenses, i.e. $\Delta z \sim 200\,\mu$m, as shown earlier in Fig. 4.14.

The refractive index of most polymer matrices is approximately 1.5 (see Table 4.2). Hence, oil immersion lenses are the most suitable for acquiring optical sections within the sample. All the images in this case study were acquired with a Nikon PlanApo $\times$60, NA 1.4, oil immersion lens.

4.7.4 Fibre tracing

The effect of fibre waviness is of most interest in the study of high performance, unidirectional composites. These materials usually contain a high fibre volume content. Therefore the maximum optical sectioning depth (which will typically

Table 4.2 Refractive indices of polymer matrices

Polymer matrix	Refractive index
PTFE	1.35
PVDF	1.42
POM	1.48
PMMA	1.49
PP	1.49
PVC	1.54–1.55
PVDC	1.60–1.63
PEEK	1.67

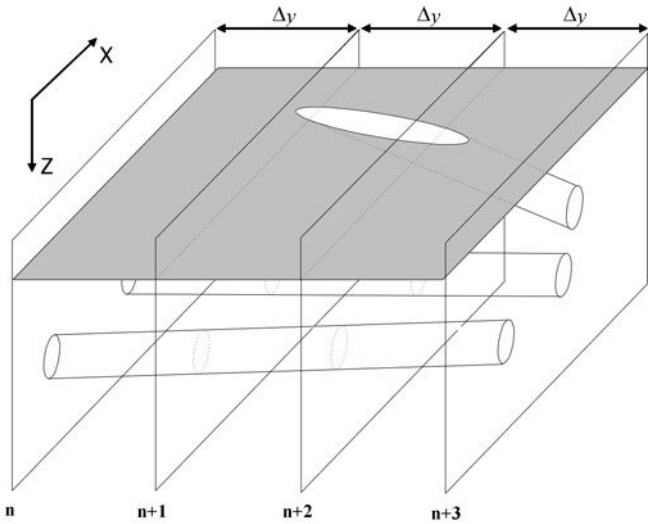


Figure 4.53 A schematic is shown of the XZ sectioning approach that is used to reconstruct fibre waviness.

be no more than 40 μm) implies that the sample should be sectioned so that fibres are oriented approximately *parallel* to the sample surface. Figure 4.53 illustrates a sampling scheme where the composite has been mounted with the fibres aligned (by eye) with the Y-axis. The fibres will appear near-circular within these XZ images and, if the image spacing is suitably small, individual fibres can be traced from each image to the next.

Early research work on fibre waviness used manual methods to trace the fibre centres over many image frames (typically 100 or more).[49] Despite its laborious nature, the results of this study provided the first accurate measurements of fibre waviness. The automated location of fibres within fluorescence mode XZ images is hindered by image noise and the intensity gradient due to light attenuation within the sample. Standard methods of segmentation by a single



Figure 4.54 A noisy XZ image which has been thresholded and segmented. The surface is at the top of the image and, due to shadowing, fibre images seem to blend into the image of a near neighbour. The scale is 100 μm $\times$ 50 μm in XZ.

intensity, even after background intensity correction, provide a poor basis for fibre identification as illustrated in Fig. 4.54. However, if the location of a fibre is known on the previous XZ image, this information can be exploited to simplify the problem of fibre location.

The following method of fibre location relies on an iterative 'search' of the image in order to find the most likely location of a fibre. The relative likelihood of a fibre with radius, r, being at a position (x, y) within an image can be determined by evaluating the intensity gradient, G, about the perimeter of the circle:

$$G = \frac{I_{In}}{I_{Out}} \qquad [4.13]$$

where I_{In} is the mean pixel intensity just inside the perimeter circle, and I_{Out} is the mean pixel intensity just outside the perimeter circle as follows:

$$I_{In} = \frac{\sum_{x,y} i_{x,y}}{n} \forall x, y \left\{ r - \Delta < \sqrt{(x - x_c)^2 + (y - y_c)^2} < r \right\}$$

$$I_{Out} = \frac{\sum_{x,y} i_{x,y}}{n} \forall x, y \left\{ r < \sqrt{(x - x_c)^2 + (y - y_c)^2} < r + \Delta \right\} \qquad [4.14]$$

where $i_{x,y}$ is the pixel intensity at the position x, y. The fibre centre is located at x_c, y_c, and the terms 'just outside' and 'just inside' are described more rigorously by the distance from the perimeter, Δ, as illustrated in Fig. 4.55. The intensity gradient function, $G(x, y, r)$, could be calculated over the entire image for all likely fibre radii. Peaks in this function would give the most probable fibre locations, although this method would be most impractical due to the number of

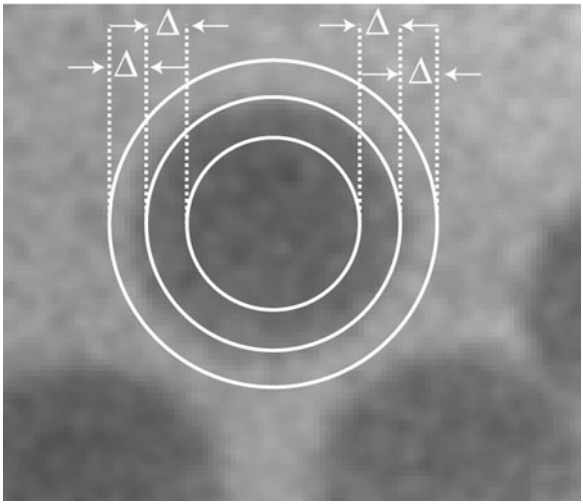


Figure 4.55 The local image intensity gradient, *G*, is computed from the intensity values within annuli of thickness, Δ, about a circle of radius, *r*.

calculations required. A more favourable method is to search for local maxima in the gradient function by using an iterative approach. This method requires the selection of a 'suitable' starting position for the algorithm, x_1, y_1, r_1. The gradient is calculated at this position and a number of points nearby, i.e. $G(x_1 + \delta, y_1, r_1)$, $G(x_1 - \delta, y_1, r_1) \ldots G(x_1, y_1, r_1 - \delta)$. If any of these neighbouring positions have a larger gradient than the initial position, they become the starting position for the algorithm on its next iteration. The iterative search for a maximum comes to a halt when no neighbouring location has a greater gradient.

In order to accelerate the convergence, the size of the perturbation, δ, is altered during the iteration process. Figure 4.56 shows the location of a fibre centre which was achieved after fifteen iterations of the algorithm with δ varying linearly from 5 pixels to 1 pixel. The position at which the algorithm comes to a halt is highly dependent on the starting position. Hence, the selection of a suitable starting point is critical. In the applications described in this chapter, it is desirable to follow a fibre over a number of optical sections, enabling the path of the fibre to be reconstructed. In this case, a suitable starting point for the fibre location algorithm is the location of the fibre in the previous image. If the centre location of each fibre moves no more than half a fibre radius from one section to the next, then each fibre may be followed accurately over the sampled volume.

4.7.5 Analysis of a multi-ply composite

A sample consisting of 32 manually assembled plies was mounted in epoxy resin and prepared using the technique described in Section 3.1.1. The sample was

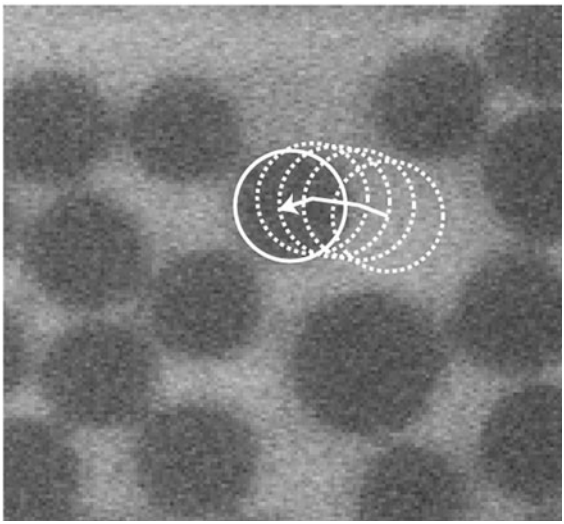


Figure 4.56 Illustration of an iterative fibre centre location search by gradually moving towards that position where the gradient is highest.

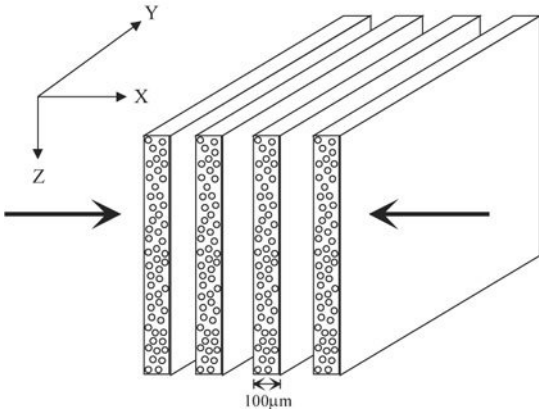


Figure 4.57 A number of plies containing both fibre and matrix are heated and compressed (in the direction of the arrows) to form the final composite. The axes illustrate the orientation at which the sample was mounted under the microscope.

placed on the microscope stage and oriented as illustrated in Fig. 4.57. The surface of the sample was scanned in reflection mode in order to calibrate the region, as described in Section 3.2.3. The calibration of the scanning stage at the sample surface enables registration in both XY and XZ images. Hence, 'extended' XZ sections may be acquired, as illustrated in Fig. 4.58, and large sample volumes may be reconstructed by combining such extended sections. In Fig. 4.59, a volume of size 440 μm $\times$ 5000 μm $\times$ 50 μm in XYZ was imaged and

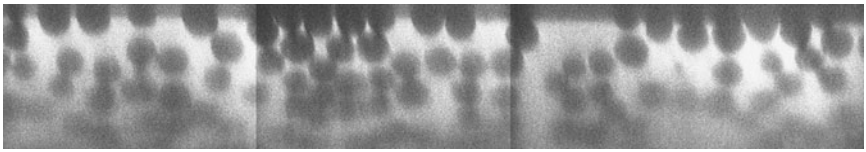


Figure 4.58 An 'extended' XZ section of a glass fibre-reinforced polymer composite, taken in fluorescence mode, using a $\times$60 NA 1.4 Nikon PlanApo objective lens. The image has dimensions of 280 μm $\times$ 50 μm in XZ.

archived as 200 extended XZ images, each separated by a distance of $\Delta y =$ 25 μm.

The fibre tracing routine described in the previous section was applied to these images in order to reconstruct the paths of fibres passing through this volume. The routine relies on the fibres having been identified on the previous image as a starting point for the iterative search. Hence, the fibres on the first extended XZ image of the dataset (or any fibres, which appear within later images) must be located by an alternative method. An interactive method was implemented which allowed the microscopist to locate the initial fibre centre locations. The reconstruction was monitored as it progressed and any obviously erroneous data were removed. With this interactive approach, it was possible to locate correctly approximately 9000 fibre centres within two hours, reconstructing the entire volume.

Figure 4.59 shows a 3D reconstruction of the sampled volume where the centre lines of the 472 fibre segments have been plotted. The aspect ratio of this image has been modified by compression in the Y direction in order to emphasise the small angle, fibre misalignments exhibited by the fibres. Despite the restricted depth penetration of 50 μm, a number of fibres have been followed over considerable distances, as revealed by the fibre segment, length distribution, shown in Fig. 4.60.

The approximate location of ply boundaries can be identified within the sampled volume, as illustrated in Fig. 4.59. The results reveal misalignments between neighbouring plies, which results in misalignments of groups (10–50) of fibres. However, misalignments of individual fibres are also apparent. Therefore, misalignment in the finished component can be introduced at the ply manufacture stage, often leading to small misalignments of a few fibres. Also, at the fabrication stage, ply misalignments can lead to gross misalignments of a large number of fibres.

The plies used to manufacture long-fibre composites are typically ribbon or sheet-like and, in this case, sheets with a thickness of approximately 100 μm were used. The thickness of this structure constrains the fibres, with misalignments more likely to be observed within the plane of the ply. This is verified by Fig. 4.61 which shows the fibre misalignment frequency plotted by projecting the fibres within the dataset onto orthogonal planes (i.e. the YZ plane and the XY plane). The out-of-plane fibre misalignment has a very tight

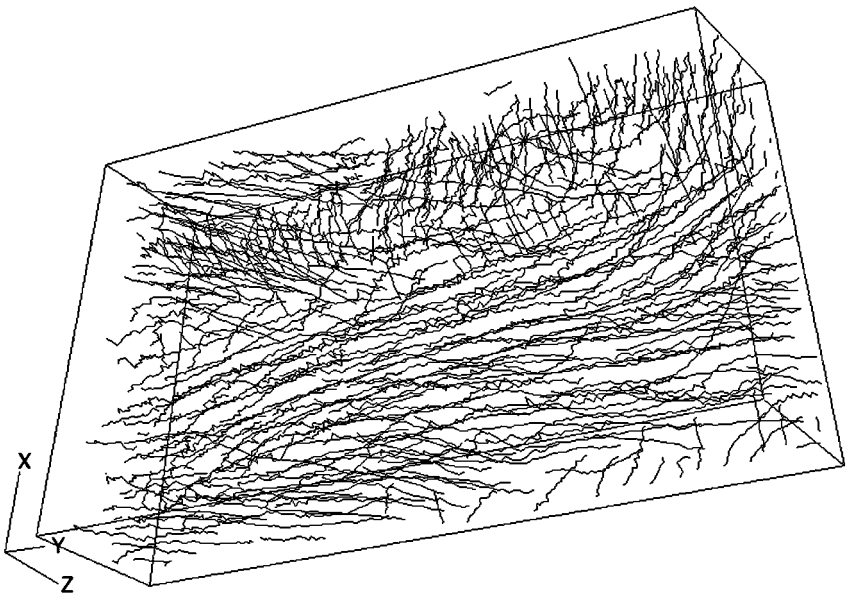


Figure 4.59 The entire sampled volume, 440 μm $\times$ 5000 μm $\times$ 50 μm in XYZ, containing 472 fibre segments is shown using the raw fibre centre data. Note that there has been a considerable compression of the illustrated volume in the Y direction to accentuate fibre waviness and fibre misalignment.

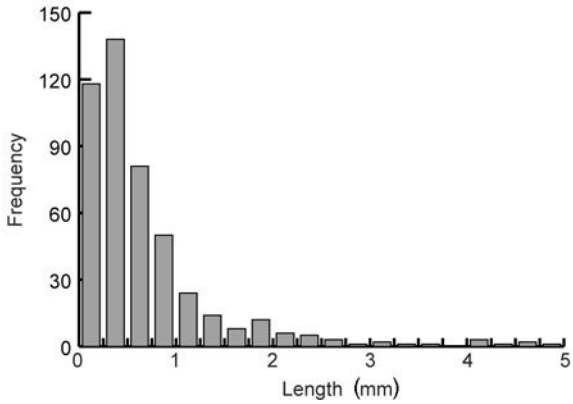


Figure 4.60 The fibre length distribution is shown for the volume reconstruction given in Fig. 4.59.

distribution with a standard deviation of 1º, as opposed to the broader distribution of in-plane data where the standard deviation is 3º. The compressive failure strength of the composite is highly sensitive to small angle fibre misalignments as seen earlier in equation [4.12]. Therefore, the magnitude of

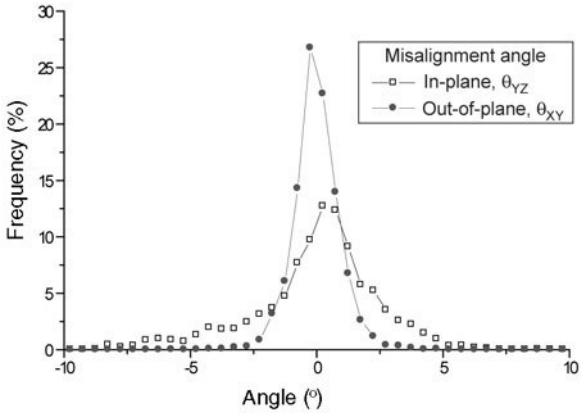


Figure 4.61 Fibre misalignment angle frequency, with fibre orientation angles calculated in the plane of the ply (in-plane) and perpendicular to the plane of the ply (out-of-plane).

fibre misalignment observed within this sample, however small it may seem, is expected to have a significant effect on the compressive strength of the sample.

4.7.6 Characterising fibre waviness

Fibre waviness is commonly characterised as a planar sinusoidal function of position,[47, 50] allowing the fibre misalignment angle to be calculated at a prescribed position within the composite. Gross fibre waviness resulting from the wrinkling of individual plies is often sinusoidal in nature. The amplitude and wavelength of the sinusoid can be measured by sectioning and inspecting the sample under a conventional light microscope. However, the waviness of individual fibres, observed in Fig. 4.59, is neither planar nor sinusoidal. The first, limited, waviness measurement by confocal microscopy was used to develop a power spectral density description of a few individual fibres.[51] These results verified the fact that the waviness of individual fibres is more complex than a simple sinusoid. However, a more detailed description of the three-dimensional movement of fibre centres throughout the sampled volume was needed in order to characterise fully the fibre waviness.

At first sight, one might think that the local fibre curvatures could be determined by simply taking the three or four adjoining fibre centre loci (from the data illustrated in Fig. 4.59) and fitting an arc through these points. However, if this method is followed, there is a high probability of measurement errors leading to spurious curvature values. Hence, the raw 3D fibre centre data must be smoothed in order to remove the intrinsic error in fibre centre location. This smoothing may be achieved by fitting high-order polynomial expressions to the raw data, effectively describing the fibre as a continuous ‘space curve’. It has

been found that there are a number of advantages to this space curve representation of fibres. Firstly, the extent of the data smoothing can be controlled by varying the order of the fitted polynomial. Secondly, the description of the fibre as a continuous mathematical function enables the calculation of fibre curvature, torsion and orientation at any point along the length of the fibre.

A space curve in three dimensions may be defined by the position vector $\boldsymbol{r}$, as follows:

$$\boldsymbol{r}(s) = f_x(s).\boldsymbol{i} + f_y(s).\boldsymbol{j} + f_z(s).\boldsymbol{k} \qquad [4.15]$$

where $f_x(s)$, $f_y(s)$ and $f_z(s)$ are polynomial functions of the arclength s, and $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are unit vectors defining the global coordinate system. The radius of curvature, $\rho(s)$ and curvature, $\kappa(s)$ of the space curve can be calculated as follows:[52]

$$\rho(s) = \frac{1}{\kappa(s)} = \frac{|\boldsymbol{r}'|^3}{|\boldsymbol{r}'.\boldsymbol{r}''|} \qquad [4.16]$$

providing a means to characterise fibre waviness. Similarly, the torsion, $\tau(s)$ of the space curve may be written

$$\tau(s) = \frac{\boldsymbol{r}'.\boldsymbol{r}''\boldsymbol{r}'''}{|\boldsymbol{r}'.\boldsymbol{r}''|^2} \qquad [4.17]$$

where $\boldsymbol{r}(s)$ is the curve's position vector, $\boldsymbol{r}'$ denotes differentiation with respect to arclength s, i.e. $d\boldsymbol{r}/ds$, and $\boldsymbol{r}''$ is the second derivative, $d^2\boldsymbol{r}/ds^2$.

A least squares method (as given by Press *et al.*[53]) was adopted for fitting polynomial functions of predetermined orders to the raw fibre data. It was found that the curvature of the fitted space curve was highly dependent on the order of the chosen polynomials, as illustrated in Fig. 4.62. The raw fibre data are plotted by projection onto the XY plane with polynomials of order 2, 4 and 8 having been fitted to the data. It is clear that the second order polynomial is a poor fit to the data, resulting in an underestimate of the curvature. The fourth order polynomial appears to be a good fit to the data, but the eighth order polynomial appears to be following the spatial noise to some extent. This clearly has an effect on the derived radius of curvature, plotted in Fig. 4.62(b). Hence, a scheme must be developed for choosing the optimum polynomial for a particular fibre segment.

Because of the lack of comparable data, it is not possible to determine which polynomial order gives the most accurate representation of the actual fibre curvature by comparing the result to existing data. In these cases, human judgement is often used to determine an intuitively 'good' fit. This is a viable method if reconstruction is being used solely for visualisation purposes.[54] However, if quantitative data are derived from this reconstruction, these data may be questionable on the grounds of repeatability. An alternative method has been

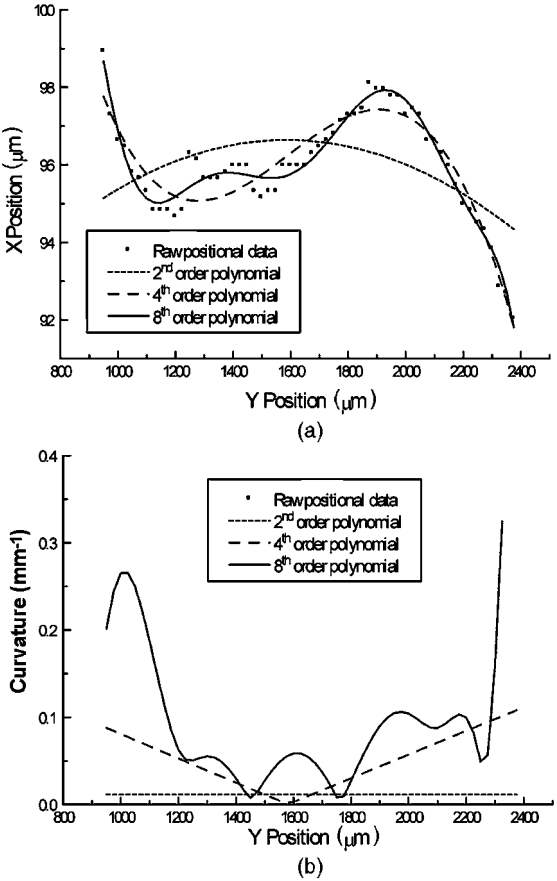


Figure 4.62 The technique of fitting polynomials of different order to the raw fibre data is shown. (a) If the appropriate order is applied, it smooths out the high frequency instrumental noise. (b) When the curvature is computed at different points along the fibre, the curvature values close to the ends of the fibre will be in error and must be ignored.

developed where fibre centre data from individual fibres of known curvature and torsion are simulated, enabling the evaluation of the fitted space curve.

In order to determine the effect of instrumental noise on the fitting of the best polynomial to the fibre centre data, it was decided to derive a 'noise function' for the CLSM measurements. The frequency distribution of measurement uncertainties was obtained by:

(a) acquiring a typical XZ frame,
(b) manually spotting the fibre centres within the image,
(c) using random seeds, instructing the automated routine to find the fibre centre coordinates, and

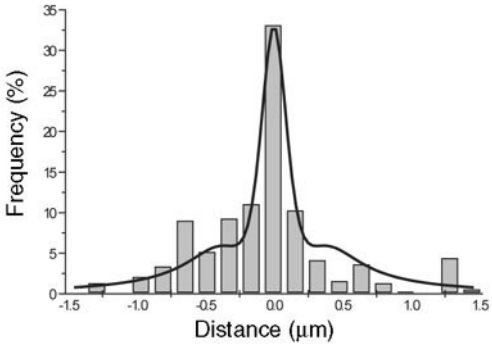


Figure 4.63 The frequency histogram of the uncertainty in fibre centre location, introduced by the automatic, fibre centre detection algorithm is shown. The error is defined as the difference between the operator-determined, fibre centre X coordinate and the software-determined, X coordinate.

(d) plotting the residual X error, δX (or the Z error, δZ) as a histogram, as shown in Fig. 4.63.

The solid curve shows the noise generation function which was used to simulate the effects of this instrumental noise and hence evaluate the effectiveness of different order polynomial fits.

The path of a fibre throughout the sampled volume may be described as a continuous function, following equation [4.15]. Using a pre-defined continuous function, a simulation of the confocal measurement technique can be implemented by applying the noise function illustrated in Fig. 4.63. The XZ sampling technique, described earlier, was simulated in this manner. Simulations of helical fibres of known curvature together with imposed noise were undertaken and the results (for a single simulated dataset) are shown in Fig. 4.64. The graph shows the sensitivity of the determination of local radius of curvature to the order of the polynomial space curve fit and, also, to the imposed value of that radius of curvature. Not surprisingly, as the fibre straightens out (as ρ increases), higher order polynomial fits follow the high frequency spatial noise and the percentage error in ρ worsens. From the graph, a polynomial order of six is optimum for fibre segments of length 2 mm (under the assumption that the fibres within this dataset exhibit curvature values within the range $0.5\,\text{mm} < \rho < 100\,\text{mm}$).

A summary of the helical simulation results is shown in Fig. 4.65. The optimum polynomial order (a polynomial order which gives an error of $|\delta\rho| < 20\%$ for helices with radii of curvature in the range $0.5\,\text{mm} < \rho < 100\,\text{mm}$) for a fibre segment of a given length is plotted. It was found that for fibre segments with a length less than 1 mm, no single polynomial order could be identified which provided a suitable fit to all the simulated helices. Therefore, it can be concluded that fibre curvature measurements derived from fibre segments with a length of less than 1 mm are unreliable.

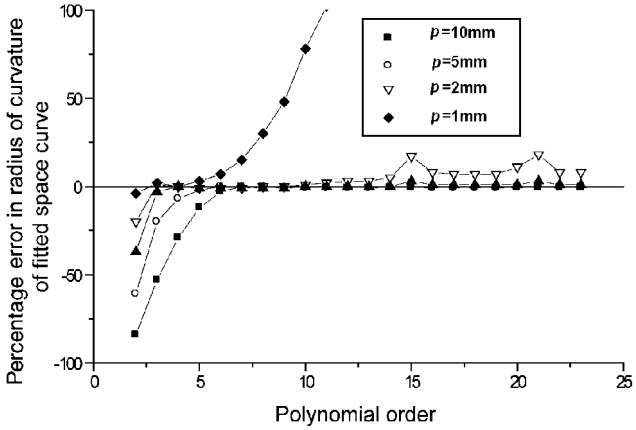


Figure 4.64 The percentage error in the mean radius of curvature of the fitted space curve for simulated fibres with a length of 2 mm.

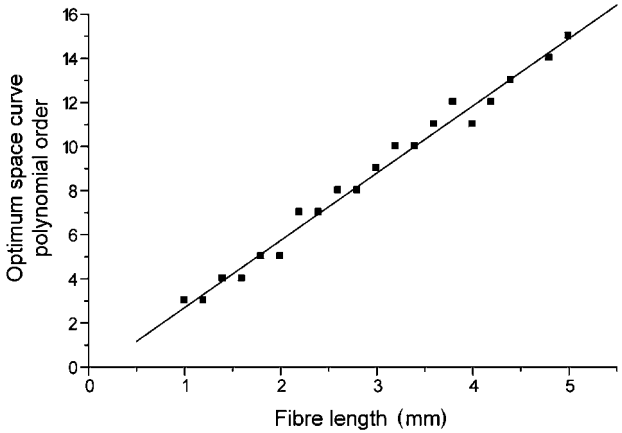


Figure 4.65 The optimum polynomial order for the fitted space curve deduced for a number of fibre segment lengths.

The polynomial fitting scheme was applied to the entire dataset and the space curve representation of the dataset can be seen in Fig. 4.66. Some fibres in the original dataset, shown in Fig. 4.59, have a length less than 1 mm and have therefore been removed. The smoothing effect of the space curve representation is quite apparent. A histogram of fibre curvature is illustrated in Fig. 4.67. This plot was created by evaluating $\rho(s)$ (equation [4.16]) at 50 μm intervals along each space curve. The radii of curvature range observed in this graph is well within the range for the simulated helices which were used to derive the space curve fitting scheme. Furthermore, an equivalent dataset (i.e.

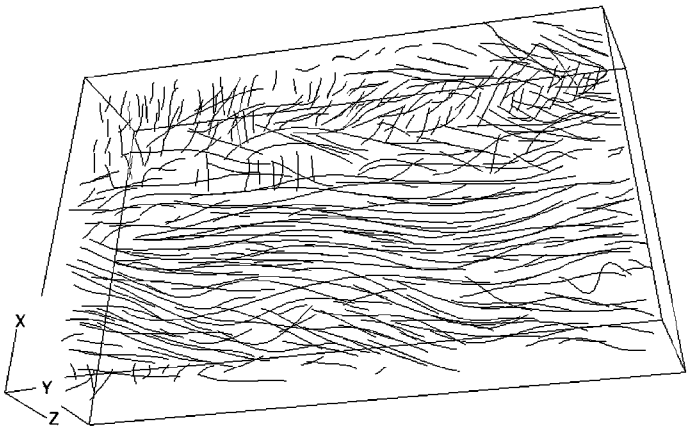


Figure 4.66 Using the polynomial fitting procedure, the raw fibre centre data, shown in Fig. 4.59, have been smoothed to give a more realistic 3D reconstruction of fibre waviness.

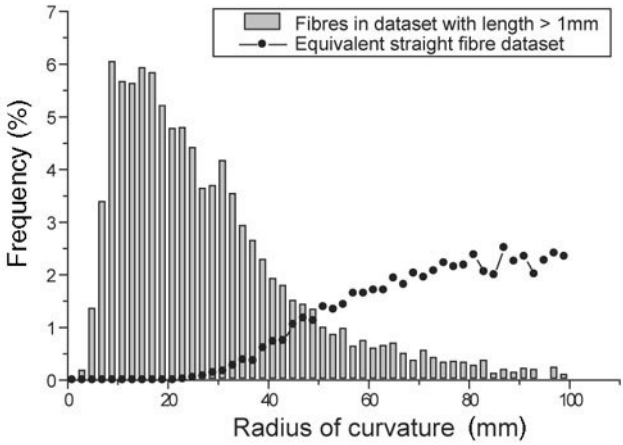


Figure 4.67 The computed radius of curvature frequency distribution for the glass fibre-reinforced polymer composite sample compared with the distribution expected if the fibres had been perfectly straight (but subject to the instrumental noise associated with the CLSM).

the same number of fibre segments with the same lengths) of straight fibres was simulated and the radius of curvature histogram plotted for comparison. This further verifies that the radius of curvature measurements made by fitting space curves to the CLSM data are due to genuine fibre curvature as opposed to instrumental noise.

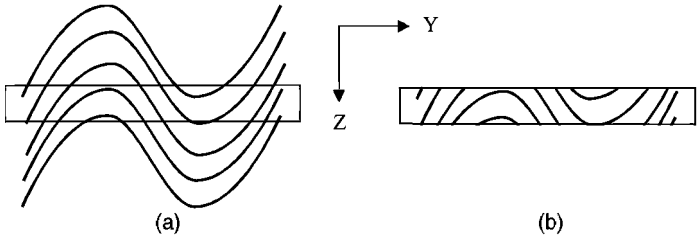


Figure 4.68 (a) Fibre waviness with an amplitude greater than the depth of the sampled volume and (b) the resulting fibre segments. Clearly, a bias in the apparent curvature may result from this effect.

Shrewd readers will have noticed that the stereological effect of the sampled volume dimensions has not been considered. The bias introduced into the determination of the fibre length distribution is illustrated in Fig. 4.68. Note that the simulation demonstrated that fibre segments of length greater than 1 mm were required to make reliable curvature measurements. However, fewer intersections of significant length will be observed when higher amplitude fibre waviness is present. Hence, the fibre length distribution is only an approximation (but we believe a reasonable one) to the real distribution.

4.7.7 Closing remarks

This case study has demonstrated how semi-automated image processing can significantly increase data productivity. The image contrast within the fluorescence mode XZ sections was of sufficient quality for a microscopist to identify immediately the fibre locations. However, fully-automated image processing routines proved less successful. Therefore, a semi-automated routine was developed which could relocate fibre cross-sections if their locations on the previous section were known. With this method, data that took a number of days to analyse by hand took just two hours by the semi-automated fibre tracing.

The sensitivity of curvature measurements to instrumental noise was seen and space curve fitting was used to ‘smooth’ the data. However, without comparable fibre curvature data it was not possible to determine by how much to smooth the data. A simulation was developed where helical fibres with instrumental noise were used to develop a fitting scheme. The resulting scheme was applied to the fibre-traced dataset, providing a radius of curvature histogram for the sample. It must be stressed that the simulation modelled the specific sampling scheme employed. If any of the parameters were changed (for example, the XZ section spacing), the simulation would have to be repeated to reflect this change, although, in future studies, the simulation technique could be used to optimise the dimensions of the sampled volumes. The simulated datasets enable prediction of measurement uncertainties for fibres of given length and section

spacing and, therefore, the sample volume required to measure a parameter to a given accuracy can be predicted.

4.8 Future possibilities for CLSM

The possibilities for confocal laser scanning microscopy in materials science seem endless. Over the past six years, the confocal technique has been used at Leeds, in many different ways, to provide insights into fibre orientation and fibre spatial distribution states, mainly in glass fibre-reinforced, polymer composites. For example, it has been used

- to propose a new method for testing the isotropy of the fibres within composites,[55]
- to remove the ambiguity in fibre orientation and thereby it has improved the understanding of particularly complex flow regions[56]
- to investigate and quantify fibre waviness and torsion, as described earlier in the case study, and enable fibre states to be characterised in novel ways, e.g. 3D nearest neighbour distances, define 3D clustering from 3D angular distributions[57]
- to establish the intrinsic errors in fibre orientation measurements with 2D image analyser systems attached to optical reflection microscopes.[58, 59]

Modifications are being made continuously to the basic confocal designs that have been discussed in this chapter. For example, in an attempt to achieve greater depth penetration, research has been initiated into optical coherence tomography (OCT), see Chapter 5. To reduce the problems of photobleaching, the signal enhancing lens system (SELS) has been proposed[60] and, also, two-photon excitation microscopy (TPE).[61]

As was mentioned in Chapter 2, there are also some unusual designs of miniaturised confocal systems that have been, or are being, built. In the past few years, a very compact and lightweight system was designed for the European Space Centre in Nordwijk for inclusion into sounding rockets.[62] However, the most intriguing design has to be the nanoscale confocal system mentioned in Chapter 1 and shown schematically in Fig. 1.71.[63] Interested readers can explore the full range of confocal microscope activity in detail in Pawley's extensive book.[5] The future does indeed look very bright for confocal microscopy!

4.9 References

1 Minsky M. (1957), Microscopy Apparatus, US Patent 3013467.
2 Egger M.D. and Petráñ M. (1967), New reflected-light microscope for viewing unstained brain and ganglion cells, *Science*, **157**, 305–307.
3 Ellis G.W. (1979), A fiber-optic phase-randomizer for microscope illumination by laser [Abstract ME1610], *Journal of Cellular Biology*, **83**, 303a.

4 Åslund N., Carlsson K., Liljeborg A. and Majlof L. (1983), PHOIBOS, a microscope scanner designed for micro-fluorometric applications, using laser induced fluorescence, *Proceedings of the Third Scandinavian Conference on Image Analysis*, Studentliteratur, Lund.

5 Pawley J.B. (1995), *Handbook of Biological Confocal Microscopy*, 2nd edn, New York and London, Plenum Press.

6 Wilson T. (1993), *Confocal Microscopy*, London, Academic Press.

7 Gee M. (1994), 'Parallel Confocal Microscopy', National Physical Laboratory, Teddington, Annual Report.

8 Brakenhoff G.J., van der Voort H., Baarslag M.W., Oud J.L., Zwart R. and van Driel R. (1988), Visualisation and analysis techniques for 3D information acquired by confocal microscopy, *Scanning Microscopy*, **2**(4), 1831–1838.

9 Knoester A. and Brakenhoff G.J. (1990), Applications of confocal microscopy in industrial solid materials: some examples and a first evaluation, *J. Microsc-Oxford*, **157**(1), 105–113.

10 Thomason J. L. and Knoester A. (1990), Application of confocal scanning optical microscopy to the study of fibre reinforced composites, *J. Mat. Sci. Letters*, **9**, 258–262.

11 Linfoot E.H. and Wolf E. (1953), Diffraction images in systems with an annular aperture, *Proceedings of the Physics Society*, **66**(B), 145–149.

12 Sheppard C.J.R. and Wilson T. (1981), Effects of high angles of convergence in V(Z) in the scanning acoustic microscope, *Appl. Phys. Lett.*, **38**, 858–859.

13 Visser T.D., Oud J.L. and Brakenhoff G.J. (1992), Refractive index and axial distance measurements in 3-D microscopy, *Optik*, **90**, 17–19.

14 Carlsson K. (1991), The influence of specimen refractive index, detector signal integration, and non-uniform scan speed on the imaging properties in confocal microscopy, *J. Microsc-Oxford*, **163**(2), 167–178.

15 Hell S., Reiner G., Cremer C. and Stelzer E.H.K. (1993), Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index, *J. Microsc-Oxford*, **169**(3), 391–405.

16 Visser T.D., Groen F.C.A. and Brakenhoff G.J. (1991), Absorption and scattering correction in fluorescence confocal microscopy, *J. Microsc-Oxford*, **163**(2), 189–200.

17 Kriete A (1992), Undesirable phenomena in 3D image cytometry. In *Visualisation in biomedical microscopies – 3D imaging and computer applications*, VCH Publishers, Chapter 9, pp. 214–218.

18 Entwistle A. and Noble M. (1994), Optimizing the performance of confocal point scanning laser microscopes over the full field of view, *J. Microsc-Oxford*, **175**(3), 238–251.

19 Sandison D.R., Williams R.M., Wells K.S., Stridder J. and Webb W.W. (1995), Quantitative fluorescence confocal laser scanning microscopy. In *Handbook of Biological Confocal Microscopy* Ed. Pawley, J.B., Plenum Press.

20 Cogswell C.J. and Larkin K.G. (1995), The specimen illumination path and its effect on image quality, Chapter 8, *Handbook of Biological Confocal Microscopy*, Ed. Pawley J.B., Plenum Press.

21 Tsien R.Y. and Waggoner A. (1995), Fluorophores for confocal microscopy: photophysics and photochemistry, Chapter 16 in *Handbook of Biological Confocal Microscopy*, Ed. Pawley J.B., Plenum Press.

22 Art, J. (1995), Photon detectors for confocal microscopy. Chapter 12 in *Handbook of Biological Confocal Microscopy*, Ed. Pawley, J.B., Plenum Press.
23 Lucas L., Gillard D. and Remion Y. (1997), A new unsupervised cube-based algorithm for iso-surface generation, *Comput. Networks ISDN*, **29**(14), 1737–1744.
24 Chan S.L. and Purisima E.O. (1998), A new tetrahedral tessellation scheme for iso-surface generation, *Computers and Graphics*, **22**, 83–90.
25 Gouraud H. (1971), Continuous shading of curved surfaces, *IEEE T. Comput.*, **20**, 623–629.
26 White N.S. (1995), Visualization systems for multidimensional CLSM images, Chapter 14 in *Handbook of Biological Confocal Microscopy*, Ed. Pawley J.B., Plenum Press.
27 Lucas L., Gilbert N., Ploton D. and Bonnet N. (1996), Visualization of volume data in confocal microscopy: comparison and improvements of volume rendering, *J. Microsc-Oxford*, **181**, 238–252.
28 Hamilton D.K. and Wilson T. (1982), Surface profile measurement using the confocal microscope, *J. Appl. Phys.*, **53**, 5320–5322.
29 Cheng P.C. and Kriete A. (1995), Image contrast in confocal light microscopy, Chapter 17 in *Handbook of Biological Confocal Microscopy*, Ed. Pawley J.B., Plenum Press.
30 Boyde A. and Jones S.J. (1995), Mapping and measuring surfaces using reflection confocal microscopy, Chapter 15 in *Handbook of Biological Confocal Microscopy*, Ed. Pawley J.B., Plenum Press.
31 www.syncroscopy.com
32 Osswald T.A. and Meges G. (1995), *Materials Science of Polymers for Engineers*, Hanser Publishers.
33 Coupland K. and Maltby A. (1997), Modification of the surface properties of polyethylene plastomer films by the use of additives, *J. Plast. Film Sheet*, **13**(2), 142–149.
34 Howard C.V. and Reed M.G. (1998), *Unbiased Stereology*, Appendix A, BIOS Scientific Publishers.
35 Cavalieri B. (1666), *Geometria degli Indivisibili*, Unione Tipografico-Editirice, Torino.
36 Howard C.V., Reid S., Baddeley A. and Boyde A. (1985), Unbiased estimation of particle density in the tandem scanning reflected light microscope, *J. Microsc-Oxford*, **138**, 203–212.
37 Karlsson L.M. and Liljeborg A. (1994), Second-order stereology for pores in translucent alumina studied by confocal scanning laser microscopy, *J. Microsc-Oxford*, **175**(3), 186–194.
38 Liljeborg A., Czader M. and Porwit A. (1995), A method to compensate for light attenuation with depth in three-dimensional DNA cytometry using a confocal scanning laser microscope, *J. Microsc-Oxford*, **177**(2), 108–114.
39 Margadant F., Leemann T. and Niederer P. (1996), A precise light attenuation correction for confocal scanning microscopy with $O(n^{4/3})$ computing time and O(n) memory requirements for *n* voxels, *J. Microsc-Oxford*, **182**(2), 121–132.
40 Argon A.S. (1972), *Fracture of Composites, Vol. 1*, Academic Press.
41 Creighton C.J. and Clyne T.W. (2000), The compressive strength of highly-aligned carbon-fibre/epoxy composites produced by pultrusion, *Comp. Sci. Tech.*, **60**, 525–533.

42 Wisnom M.R. (1994), The effect of fibre waviness on the relationship between compressive and flexural strengths of unidirectional composites, *Journal of Composite Materials*, **28**(1), 66–76.

43 Mrse A.M. and Piggott M.R. (1993), Compressive properties of unidirectional carbon fibre laminates: II. The effects of unintentional and intentional fibre misalignments, *Comp. Sci. Tech.*, **46**, 219–227.

44 Yurgartis S.W. (1987), Measurement of small angle fiber misalignments in continuous fiber composites, *Comp. Sci. Tech.*, **30**(4), 279–293.

45 Clarke A.R., Davidson N.C. and Archenhold G. (1997), Mesostructural characterisation of aligned fibre composites, Chapter 7 in *Flow-induced Alignment in Composite Materials*, Woodhead Publishing Ltd, Cambridge.

46 Joyce P.J., Kugler D. and Moon T.J. (1997), A technique for characterizing process-induced fiber waviness in unidirectional composite laminates – using optical microscopy, *J. Compos. Mater.*, **31**(17), 1694–1734.

47 Highsmith A.L., Davis J.J. and Helms K.L.E. (1992), The influence of fiber waviness on the compressive behaviour of unidirectional continuous fiber composites, *Composite Materials: Testing and Design (Tenth Volume)*, ASTM STP 1120. American Society for Testing Materials, Philadelphia, pp. 20–36.

48 Paluch B. (1996), Analysis of geometric imperfections in unidirectional reinforced composites, *J. Compos. Mater.*, **30**(4), 454–485.

49 Clarke A.R., Archenhold G. and Davidson N.C. (1998), 3D Confocal microscopy of glass fibre reinforced composites, Chapter 3 in *Microstructural Characterisation of Fibre-reinforced Composites*, Woodhead Publishing Ltd, Cambridge.

50 Hsiao H.M. and Daniel I.M. (1996), Elastic properties of composites with fibre waviness, *Comp. Sci. Tech.*, **56**, 581–593.

51 Clarke A.R., Archenhold G., Davidson N.C., Slaughter W.S. and Fleck N.A. (1995), Determining the power spectral density of the waviness of unidirectional glass fibres in polymer composites, *Appl. Comp. Mater.*, **2**, 233–243.

52 Nutbourne A.W. and Martin R.R. (1998), *Differential Geometry Applied to Curve and Surface Design*, Ellis Horwood Limited.

53 Press W.H., Teukolsky S.A., Vetterling W.T. and Flannery B.P. (1996), *Numerical Recipes in C, 'The Art of Scientific Computing'*, Cambridge University Press.

54 Shelden E. and Knecht D.A. (1997), Reconstruction and display of curvilinear objects from optical section data using 3-D curve fitting algorithms, *J. Microsc-Oxford*, **191**(1), 97–107.

55 Mattfeldt T., Clarke A.R. and Archenhold G. (1994), Estimation of the directional distribution of spatial fibre processes using stereology and confocal scanning laser microscopy, *J. Microsc-Oxford*, **173**(2), 87–101.

56 Clarke A.R., Eberhardt C.N. and Davidson N.C. (1999), 3D characterisation of glass fibres in composites by confocal microscopy, *Proceedings of the 12th Int. Conf. on Composite Materials (ICCM12)*, Paris, p. 471 (extended abstracts), paper 167.

57 Clarke A.R. and Eberhardt C.N. (1999), The representation of reinforcing fibres in composites as 3D space curves, *Comp. Sci. Tech.*, **59**, 1227–1237.

58 Eberhardt C.N. and Clarke A.R. (2001), Fibre orientation measurements in short glass fibre composites. Part I: automated, high-angular resolution measurement by confocal microscopy, *Comp. Sci. Tech.*, **61**, 1389–1400.

59 Eberhardt C.N., Clarke A.R., Vincent M., Giroud T. and Fluoret S. (2001), Fibre orientation measurements in short glass fibre composites. Part II: a quantitative error

estimate of the 2D image analysis technique, *Comp. Sci. Tech.*, **61**, 1961–1974.
60 Reichelt S. and Amos W.B. (2001), SELS: A new method for laser scanning microscopy of live cells, *Microscopy & Analysis*, November, 9–11.
61 Diaspro A., Corosu M., Ramoino P. and Robello M. (2000), A two photon excitation microscope based on a compact confocal scanning head, *Microscopy & Analysis*, July, 21–24.
62 Schiller P. (1994), Confocal laser scanning microscope, *European Space Agency Bulletin*, Summer, 8–9.
63 Zandonella C. (2001), Sick cells can't escape the eagle eye of the micro-microscope, *New Scientist*, **172** (2320), 24.

Part III

Other microscopical techniques

5

Complementary optical and EM imaging techniques

'Concerning the nature of light … use the corpuscular theory on Monday, Wednesday and Friday and the undulatory theory on Tuesday, Thursday and Saturday'

… attributed to Sir William Henry Bragg (1928)

5.1 Introduction

In this chapter, some of the alternative techniques to optical reflection microscopy and confocal laser scanning microscopy will be discussed, with an emphasis on those measurement techniques which are based on optical and other types of electromagnetic (EM) radiation.

As discussed in Chapter 1, the same basic equations apply to every different type of EM radiation. The key difference is the effective wavelength of the radiation. Novel fabrication processes are beginning to hold the promise of nanometre scale size machines and instruments and one of the main reasons for exploring non-optical microscopies is to achieve smaller spatial resolutions – into the nanometre region. A number of authors have suggested how to beat the conventional diffraction limit for optical microscopy (see, for example, Danuser,[1] Gustafsson[2] and Toraldo di Francia[3]), by revisiting the ideas behind information capacity and imaging systems. For example, the traditional Rayleigh criterion for determining spatial resolution will be improved upon if computer curve fitting to the intensity distributions of two close point sources is attempted. Rayleigh's criterion was just a qualitative rule of thumb but the curve fitting is more quantitative. However, the way to achieve higher spatial resolution is to use either lower wavelength radiation (X-rays) or the latest scanning probe microscopes. Also, by using more energetic, non-optical radiation, greater depths of penetration through most materials become possible too.

When Sir William Henry Bragg was professor of physics at the University of Leeds from 1909 through to 1915,[4] together with his son, Sir William Lawrence Bragg, he initiated the research field of X-ray crystallography. The Bragg equation, $n\lambda = 2d\sin\theta$, which linked the X-ray wavelength, λ, with the glancing

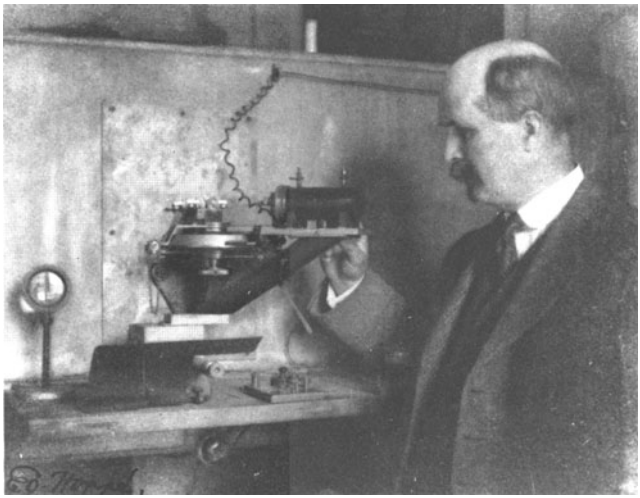


Figure 5.1 A photograph taken around 1915 showing Sir William Henry Bragg with his early X-ray crystallography experimental rig.

angle, θ, for strong reflections off crystal structure, is the cornerstone of X-ray crystallography. They realised that they could use X-rays to probe the 3D structure of materials with high spatial resolution. Their original X-ray spectrometer, built in 1915, could be described as a 3D microscope of sorts (see Fig. 5.1). A commercially-built, Bragg X-ray crystallography apparatus that was used extensively at Leeds is shown in Fig. 5.2. The latest table-top, X-ray microtomography device for probing material structures will be discussed in detail later in this chapter.

However, before exploring the capabilities of X-ray systems, Raman microscopy and scanning probe microscopy will be reviewed. The optical microscopy scene is as active as ever, with each new type of microscope design being further modified to incorporate many variations on a theme. Indeed the trend nowadays seems to be multi-modal microscopy where different techniques are investigated within the same apparatus or where a basic design may have a number of front-end sensing attachments, which allow the basic (expensive) design to be more versatile.

The insertion of scanning probe microscopies into this chapter may seem strange, because one could argue that it is not really an EM technique. Our justification for placing it here is twofold. Firstly, it allows the introduction of the principles underlying various near-field optical methods, and secondly, with many of the original designs, laser light was used to determine the movement of the probe tip and hence, indirectly, determine the spatial resolution via the optical cantilever principle.

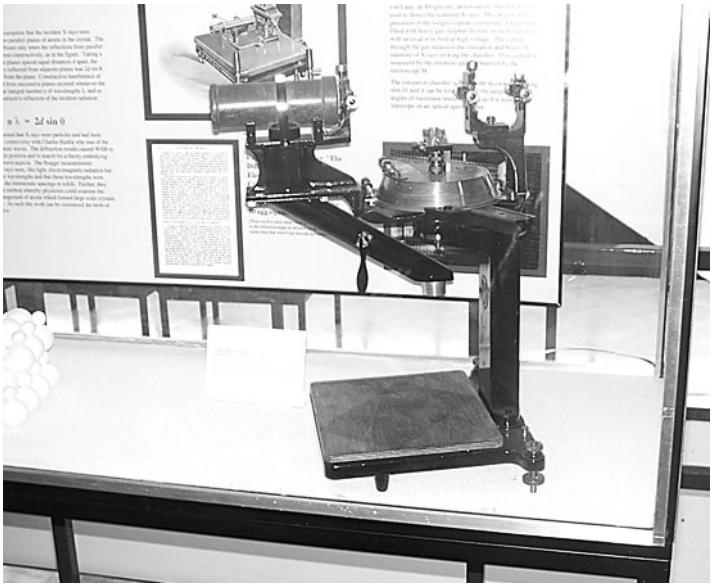


Figure 5.2 A commercially-built spectroscope (PYE, Cambridge) used for X-ray crystallography studies over many years at Leeds.

5.2 Raman microscopy

The Raman effect was first observed by C.V. Raman and K.S. Krishnan in 1928,[5] although it was first predicted by Smekal[6] five years earlier. It was found that, if monochromatic radiation, I, is incident on an optically perfect transparent liquid, a very small fraction of the radiation ($10^{-8}I$) is scattered by the sample. This scattered light contains the frequency of the incoming radiation as well as pairs of new frequencies, which are higher and lower than the incident radiation frequency. Different molecular deformational modes are responsible for either infrared or Raman emission lines (see Fig. 5.3). The simplest way to understand these extra frequencies is by considering the polarisation, P, of the material, which depends on the polarisability, α, of the material and on the electric field, E, of the incident light at a frequency, ν_0.[7]

$$P = \alpha E_0 \cos(2\pi\nu_0 t) \qquad [5.1]$$

Charges in the material may oscillate at a frequency, ν_m, about their equilibrium positions:

$$q = q_0 \cos(2\pi\nu_m t) \qquad [5.2]$$

Usually, the electric field magnitude, E_0, is presumed to be constant over the dimensions of the oscillations (which is an excellent approximation for visible light in far-field configurations). The oscillations may, however, induce a

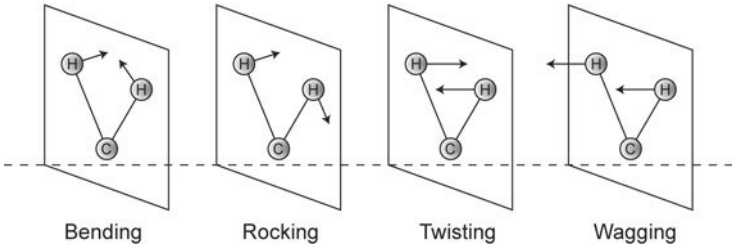


Figure 5.3 Infrared and Raman Spectroscopy give information on molecular vibrations. Shown schematically are the four deformational modes of the CH_2 molecule, which give rise to infrared and Raman lines: wagging, twisting, rocking and bending.

change in the polarisability of the material and this change may be calculated using a Taylor series expansion of α:

$$\alpha = \alpha_0 + \left(\frac{\partial \alpha}{\partial q}\right) q + O(q^2) \qquad [5.3]$$

where $O(q^2)$ denotes a term 'of the order of q^2'. As this q^2 term is much smaller than the q term, it may be ignored, together with other terms in the series. Therefore, combining these equations and simplifying them gives, to first order in q,

$$P = \alpha_0 E_0 \cos(2\pi\nu_0 t) + \frac{1}{2} q_0 \left(\frac{\partial \alpha}{\partial q}\right)_0 E_0 [\cos\{2\pi(\nu_0 - \nu_m)t\} + \cos\{2\pi(\nu_0 + \nu_m)t\}] \qquad [5.4]$$

The first term is simply a mathematical description of Rayleigh scattering and the second term describes the Raman scattering effect. Note that the existence of Raman-scattered light depends on the Raman activity, $\delta\alpha/\delta q$. If there is no change in the polarisability for a given vibrational transition, then $\delta\alpha/\delta q = 0$ and the transition is not Raman-active. In 3D, it is the change in the polarisability tensor that determines the activity, and the vibration is Raman-active if any one of the polarisability tensor components is changed during the vibration. Raman-scattered light is therefore detected at frequencies given by $(\nu_0 - \nu_m)$ and $(\nu_0 + \nu_m)$.

The relationship between the Stokes and the anti-Stokes Raman signals is shown in Fig. 5.4. Note that it is customary to give the Raman spectra in terms of the wave number measured with respect to the wave number of the exciting laser line taken as zero. The wave number of a Raman line is therefore given by $1/\lambda\,\mathrm{cm}^{-1}$ where $1/\lambda = 1/\lambda_1 - 1/\lambda_2$ and both the Raman line wavelength, λ_1, and the exciting source wavelength, λ_2, are in cms. The Stokes lines appear at lower frequencies than the exciting radiation whereas the anti-Stokes lines appear symmetrically at higher frequencies than the exciting radiation. Hence,

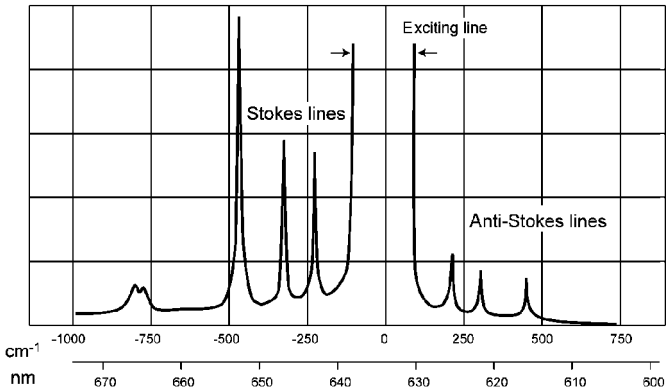


Figure 5.4 A typical Raman intensity spectral plot of liquid carbon tetrachloride showing both Stokes and anti-Stokes spectral lines symmetrically placed about the He-Ne laser exciting line at 632.8 nm. The intensity of the lines is on an arbitrary scale, but note the weakness of the Raman lines with respect to the excitation line intensity. Typically the Raman line intensity is 10^{-8} times the incident line intensity.

for the anti-Stokes lines to occur, the incident photon must have had extra energy imparted (taken from the thermal energy of the scattering molecules). Hence anti-Stokes lines vanish at absolute zero (-273ºK) where there would be no molecular thermal energy to impart.

Raman spectroscopy was revolutionised in the early 1960s with the development of continuous wave visible lasers, which enabled Raman spectra to be acquired from small volume samples. More recently, improvements in optical filter designs and more sensitive light detectors (like charge coupled devices, CCDs, which were discussed in Chapter 2) have greatly improved the application of the Raman effect to microscopy.

Note that, although it is easier to imagine the classical description of the Raman processes, all of these 'explanations' have an alternative quantum mechanical and mathematical description. One can equally well talk about 'transitions in vibration levels due to coupling with a radiation field which are described by the perturbation Hamiltonian'. Interested readers should see, for instance, Ayars *et al.*[8]

5.2.1 The 'Ramanscope'

Members of the Molecular Physics and Instrumentation Research Group at Leeds have been involved in the development of the Renishaw Ramanscope (Williams *et al.*[9]) since 1994. This benchtop instrument is shown in Fig. 5.5 and the design was created around attaining Raman images with a spatial resolution of 1 μm and a spectral resolution ($1/\lambda$) of 1 cm^{-1}. A schematic overview of the optical train within the Ramanscope is shown in Fig. 5.6. The design is

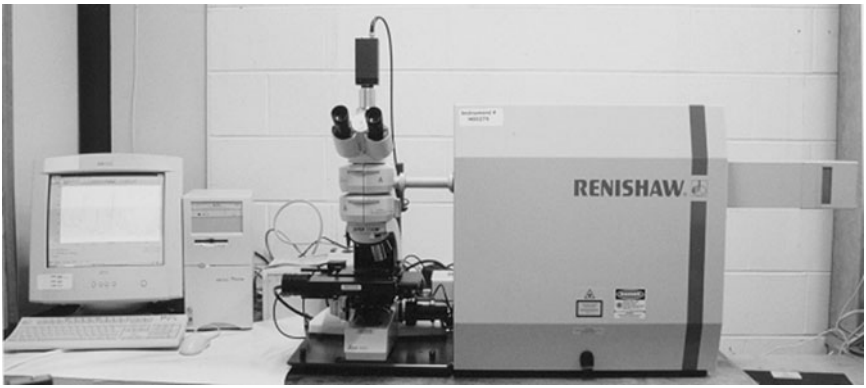


Figure 5.5 The Ramanscope™ – a photograph of the Renishaw Raman system attached to a standard optical microscope.

discussed in detail in Baldwin *et al.*,[10] but the salient features are mentioned briefly below.

A neutral density filter wheel selects different intensities of laser light to illuminate the specimen. The laser beam goes through a beam expander and spatial filter, reflects off two mirrors, M_1 and M_2, and is directed towards the first notch filter, NF_1. The small (10º) angle of incidence ensures that the polarisation properties of both the reflected laser beam and the transmitted Raman-scattered light from the specimen are not strongly affected, and also that 98% of the incident laser beam is specularly reflected. The Raman scattered light that is transmitted by the holographic notch filter then passes through another dielectric edge filter, NF_2, which suppresses further the Rayleigh scattered light (at the same frequency as the incident laser light). With this arrangement, at Raman wavenumbers greater than 100 cm^{-1}, the transmission exceeds 90% with little wavelength variation. A slit placed at the point of confocality acts as a spatial filter allowing 3D confocal operation to be performed. The Raman light then passes through the dispersing element, which is a diffraction grating and reflecting prism combination. Finally, the light beam is imaged by the tube lens, L_2, and a charge-coupled device (CCD) detector (which may be Peltier-cooled for added sensitivity) records the light intensity at different wavelengths. A typical spectral scan, which illustrates the spectral resolution achievable, is shown in Fig. 5.7(a). This interesting plot shows data taken from a conjugated polymer material, which has been subjected to high pressure in a diamond anvil cell (see Baldwin *et al.*[10]). One of the effects of the application of pressure to the resonance Raman spectrum of this material is shown in Fig. 5.7(b). The Raman band investigated corresponds to a bond-bending mode with large amplitude of vibration about the triple bond on the backbone of the material (called poly[bis(*p*-toluene sulphonate) of 2,4-hexadiyene-1,6-diol] or PDA-TS for short). It can be seen that the Raman band shifts in frequency and splits into two peaks at 50 kbar. Figure 5.7(b) shows both

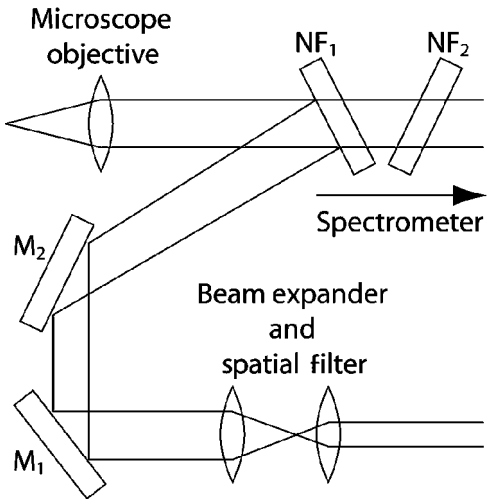


Figure 5.6 The optical train within the Ramanscope is shown, illustrating how the laser beam can be introduced into the microscope and the Rayleigh light can be rejected from the spectrometer. The elements NF_1 and NF_2 are notch or edge filters that reject the laser light. Note that NF_1 also acts like a mirror, like M_1 and M_2 at the laser wavelength.

the relative Raman shifts of these peaks and also the separation between these peaks as the pressure increases. These data illustrate how the technique of diamond anvil cell pressurisation can reveal fundamental phenomena that might otherwise be hidden at lower hydrostatic pressures. Various adaptations of the Ramanscope are being researched and these are discussed below.

5.2.2 Other Raman-related techniques

When the exciting radiation falls within the frequency range of a molecule's absorption band in the visible or ultraviolet range, the radiation may be scattered by two different processes – the resonance fluorescence or resonance Raman effect. Both processes give much more intense scattering (resonance Raman intensity peaks may be 100 to 1000 times greater than normal Raman peaks). The resonance Raman effect was anticipated by Placzek (in 1934) in some pioneering work on the polarisability theory of Raman scattering. By simply altering the direction of a polariser within the optical train, so that it passes Raman radiation, which is perpendicularly polarised to the excitation laser light, these enhanced Raman peaks can be obtained.

The stimulated Raman effect depends on the coherent pumping action of the molecules of the sample into an excited vibrational state by the powerful electric field of the laser beam. It was first observed by Woodbury and Ng in 1962 in liquid nitrobenzene, when they were investigating its use as an electro-optical

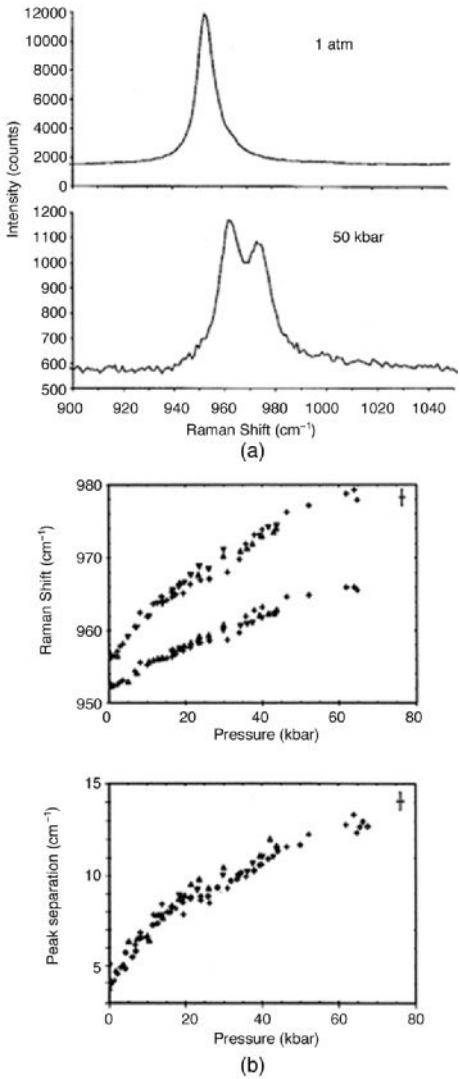


Figure 5.7 Typical spectral scans to illustrate the resolution achievable with the Ramanscope. (a) Part of the resonance Raman spectrum of the conjugated polymer PDA-TS at atmospheric pressure and at a hydrostatic pressure of 50 kbar. (b) The centre wavenumber of the two components of the 952 cm^{-1} Raman band of PDA-TS is shown as a function of applied pressure and their separation as a function of pressure.

shutter.[11] One interesting technique, which depends on the stimulated Raman effect, is called the coherent anti-Stokes Raman spectroscopy (CARS).[12] In this technique, two lasers are used where one laser has a fixed frequency and the other laser has a tunable frequency. The two beams enter the sample at angles

differing only by a small amount (approximately 2º) and simultaneously impinge on the sample molecules. Whenever the frequency difference between the two lasers coincides with the frequency of a Raman-active vibration of the molecules, emission of coherent radiation (both Stokes and anti-Stokes) is stimulated. Hence, the total Raman spectrum can be scanned in stimulated emission by varying the frequency of the tunable laser. An advantage of CARS is that, in addition to the high level of scattering achieved, the elevated laser frequencies avoid interference from the fluorescence (which always has frequencies below that of the exciting radiation).

Other Raman-related techniques, like Fourier Transform Raman for example, can be found in Turrell and Corset[12] as well as a combined confocal and scanning near-field Raman microscope configuration for ultraviolet laser excitation, which is discussed in the next section. Another curious effect that has been explored over the past 20 years is Surface-Enhanced Raman Spectroscopy (SERS) (see an excellent review by Moskovits[13]). SERS refers to the observation that, for certain molecules adsorbed on specially prepared metal surfaces, a Raman spectrum is observed whose intensity exceeds by a factor of $10^5 - 10^6$ what one would expect on the basis of simple calculations! This discovery has made physicists revisit the theories of electrochemistry, charge transfer and chemical aspects of adsorption and has stimulated surface science.

Raman spectroscopy is of considerable value in determining molecular structure and in chemical analysis. Molecular rotational and vibrational frequencies can be determined directly. An analysis of these frequencies sometimes enables an evaluation of the molecular geometry (or at least to find the molecular symmetry). Even when a precise determination of structure is not possible, much can be learnt about the arrangement of the atoms in a molecule from empirical information about the characteristic Raman frequencies of groups of atoms. Raman and infrared spectra often provide complementary data about molecular structure. The complex structures of biologically important molecules are the topic of current spectroscopic research. Raman spectra also provide information to condensed matter physicists about lattice dynamics and the electronic structures of solids.

Using ever more sensitive detectors and better filters, Raman spectra can be obtained from minute samples. Hence it finds application in the determination of chemical species within the very small samples in forensic science research (see, for example, Betzig *et al.*[14]), and also in semiconductor research (Goetz *et al.*[15]). The ability to measure localised stress within materials has opened up new opportunities in composites research. Young[16, 17] has published extensively over the past few years on the applications of the Raman spectroscopic technique to aramid and carbon fibre-reinforced composites. The typical spectral shift in a fibre's Raman peak due to an applied strain is shown in Fig. 5.8(a) and the linearity between applied strain and spectral shift in Fig. 5.8(b). By mapping these stress-induced, Raman band shifts in the fibres, the relationship between structure and deformation processes can be investigated (and also see Galiotis[18]).

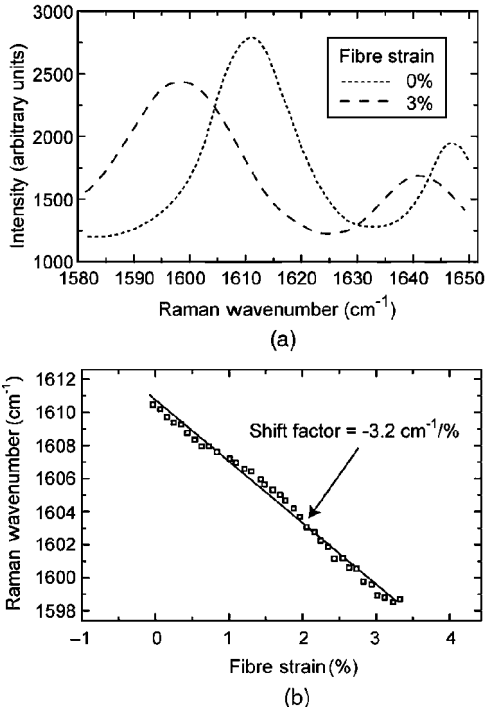


Figure 5.8 The Raman effect has been used by Young[16] to determine the variation of strain within a fibre. (a) The band shift of a typical Raman line from 1666 cm^{-1} to 1600 cm^{-1} on stressing the fibre to a strain of 3%. (b) Data points illustrating the linear shift in Raman line with increasing strain.

5.3 Scanning probe microscopy/near-field microscopy

In the early 1980s, a totally new experimental microscopy technique was announced which showed stunning images of individual atoms. Many aspects of materials science now require the quantification of specific properties of, and interactions between, materials. Polymer scientists may want to measure physical characteristics like adhesion, compliance and friction, and thermal properties like conductivity, expansion, phase transition temperature and calorimetry of individual components in blended or composite polymers. Also, semiconductor engineers may need to measure the physical dimensions of nanoscale structures or electrical properties like capacitance, conductivity and potential. Biologists may want to measure the binding forces between a specific bio-molecule and its receptor site on a cell surface or between antigens and antibodies, or to observe single molecule fluorescence. The scanning probe microscope (SPM) design, first articulated by Binnig *et al.* in 1982,[19] was followed by the scanning force microscope (SFM or atomic force microscope –

AFM) in 1986.[20] These designs have set the scene for many variations on a theme, all designs allowing the 3D imaging of sample surfaces with spatial resolutions down to nanometre scale sizes.

Like all good ideas, the concept behind these devices is very simple. Indeed, it is reminiscent of Edison's technique for the first phonograph or the first profilometers, which arranged for a stylus to be dragged over a surface and the lateral motion of the stylus to be sensed by some appropriate means. Instead of a stylus, a special, tapered microprobe is used and piezoelectric transducers control the motion of this probe. The operation of an SPM consists of the mapping of specific surface characteristics through an analysis of the interaction between a microprobe and the sample surface. There is an expanding range of microprobes that can be attached to the basic configuration and each probe determines the type of microscope, e.g. magnetic force microscope (MFM), electrostatic force microscope (EFM), scanning thermal microscope (SThM), pulsed force microscope (PFM), force modulation microscope (FMM) and scanning capacitance microscope (SCM).

Different SPM scanning modes can be classified by the forces that they monitor and the distance between the tip and the sample surface over which the forces operate, as shown in the force–distance curve in Fig. 5.9. There are three possible operational modes with these devices: contact, intermittent contact or non-contact modes. Certain measurements such as friction, compliance, adhesion and thermal characteristics require contact with the sample, so contact mode operation must be used. Whenever possible, however, contact mode is avoided because of the inherent dangers of damaging the sensitive probe tip (or the sample surface itself).

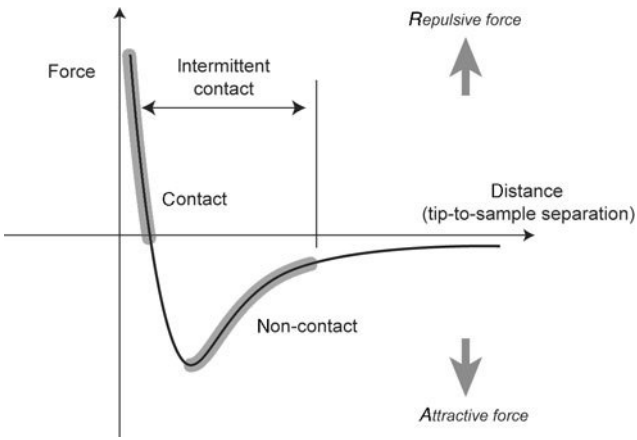


Figure 5.9 A typical force–distance curve experienced by a probe tip as it approaches a surface. The regions of interest for the different SPM techniques like tapping mode and non-contact modes are shown.

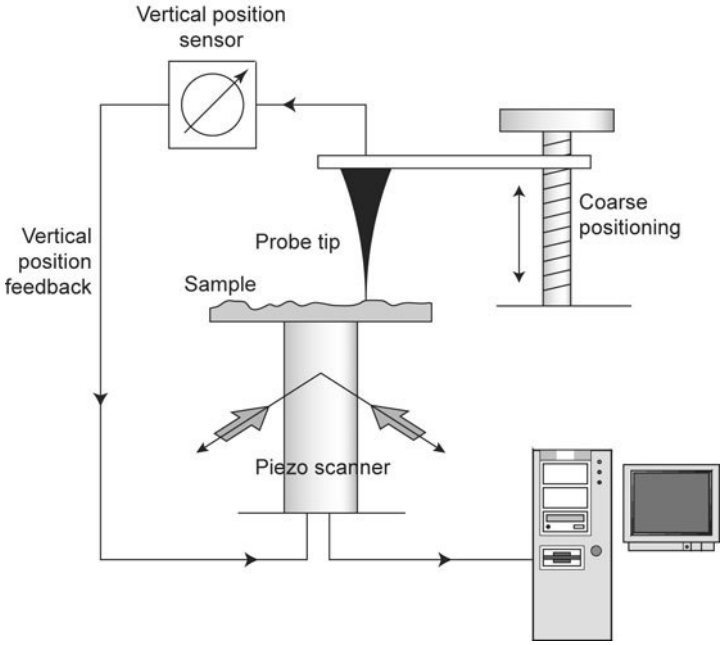


Figure 5.10 A schematic overview of all SPM techniques is illustrated. There will be some means of determining the location of the probe tip relative to the sample surface. The sample is moved in X, Y and Z by piezoelectric devices for the final positioning and scanning.

A schematic overview of all SPM techniques is shown in Fig. 5.10. Typical separations, established between the probe tip and the surface, are 50–200 nanometres. Therefore an effective anti-vibration mounting is always essential and also a progressively fine positioning system to place the probe tip close to the surface. Piezoelectric scanners are required to move either the probe or the sample to area scan in X-Y. A means of sensing the probe tip's vertical position (relative to the sample surface) and, in non-contact operational mode, a feedback loop whose function is to keep the probe-tip at a fixed distance from the surface, is essential. The combination of the X-Y scan position data and the feedback signal gives the raw data for a 3D topological image of the sample surface.

The characteristic mapped by the contact-mode SFM is the repulsive force exerted by each point of the sample against a sharp tip set at the end of a soft micro-fabricated cantilever. For stiff materials this repulsive force corresponds to the sample topography. However, non-contact or intermittent contact operation helps to prevent or reduce the mechanical tip-sample interaction,[21, 22] whereas force modulation techniques allow us to map the mechanical characteristics of the sample and to infer sample shape at zero contact force. Hence, it is possible to build a 3D image of the surface topography of practically any material. Even liquid phases or droplets, either absorbed or deposited on a

surface, can be imaged, without significantly perturbing them, by special force techniques.[23]

All commercially available SPMs use piezoelectric scanners to move the probe tip relative to the sample. Piezoelectric materials extend or retract in response to an applied electric potential. Although they provide fine movement control (nanometres/volt excitation), these scanners exhibit several well-known non-linearities in their response (creep, hysteresis and various historical effects, for example). Some of these errors can be modelled and predictively-corrected, but others are inherently unpredictable. Therefore accurate control of the probe tip motion requires active position sensing and closed-loop feedback to the control system.[24] All spatial measurements in an SPM presume the accuracy of the scan. Any scanning error contributes directly to measurement error and therefore active closed-loop scan correction is vital. On some systems, this scan correction is based on optical or capacitance position sensing.

Scanning thermal microscopy uses a heated probe and measures the electrical current required to maintain the probe at a constant elevated temperature as it scans the surface. Changes in local thermal conductivity cause the fluctuations in the controlling heating current. Microthermal analysis uses the same heated probe to derive calorimetric information from a sample volume of a few cubic microns by plotting the amount of heat required to raise the sample temperature. Hence this technique can measure phase-transition temperatures and identify a material by its unique calorimetric fingerprint.

Measurements made in a vacuum can pose a problem for the laser light lever approach, and piezo-cantilevers, where deflections generate an electrical signal in the cantilever, but they do provide a viable alternative for vacuum work. Near-contact mode of operation can provide quantitative measurements of weak interactions operating over short distances. For these systems, special soft cantilevers are required for the high frequency, low amplitude probe modulation. The signal detection scheme must also have phase shift sensitivity to enhance the contrast of the images created.

5.3.1 STM

The mechanism of the STM is illustrated in Fig. 5.11. If a pointed metal probe is placed sufficiently close to a sample and a small voltage (say ΔV about 10 millivolts) is applied between the probe and the sample, then electron tunnelling can occur. A quantum mechanical calculation gives an approximate expression for the relationship between the tunnelling current, I, and the sample-tip separation, s:

$$I = ke^{-As\sqrt{\zeta}} \qquad [5.5]$$

where the constant $A = 10.25\,\mathrm{eV}^{-1}\mathrm{nm}^{-1}$, ζ is the average barrier height between the two electrodes and k contains information about the density of states

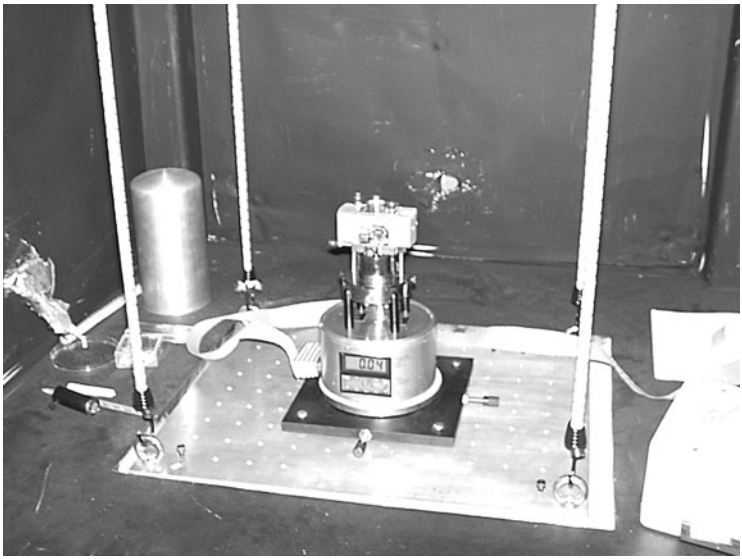


Figure 5.11 Photograph of a typical Scanning Tunnelling Microscope (STM) used in Leeds, showing the supporting rig, which minimises mechanical vibrations.

of the tip and the sample. The net flow of electrons can be measured by this tunnelling current, which is proportional to the transmission probability. If the probe is then scanned across the sample, any bumps on the surface will change the separation of probe and surface and, because of the exponential relationship between separation and tunnelling current, separation changes of the order of 0.01 nm may be measured. Hence this technique can generate a topographical map of the surface.

Experimentally, there are a number of technical difficulties, which were first overcome by Binnig.[19] Vibrations of the apparatus must be minimised because typical separations of the probe and surface might be a few nanometres, whereas floor vibrations could be as large as 1 μm. Therefore the designs must isolate the specimen–probe area from external disturbances. The second problem is probe sharpness, which must be as sharp as possible to give high spatial resolution. Metal wires are polished electrochemically and, although the wire is rough on an atomic scale, the tunnelling effect will occur at one of the mini-tip regions on the end of the wire, which is probably a few atoms thick at its extremity. A third problem is position control of the probe tip. The development of piezoelectric ceramic position controllers has solved this problem. The piezoelectric material expands or contracts as an external voltage is applied to electrodes on opposite faces, and probe tip movements of a few tenths of a nanometre per applied volt are possible. Hence, a practical STM consists of a number of position controls – from the finest control by piezoelectric arms through to coarser control by

stepper motor and also by hand. Tunnelling microscopes operate in vacuum, air, liquid helium, oil, water and even in electrolytic solutions.

5.3.2 AFM

A classic design for an atomic force microscope (AFM) is shown in Fig. 5.12. The light from a laser diode reflects off the cantilever probe and is detected by a quadrant photodiode array. This is the most widely employed method. The quadrant photodiode signals can be processed to give information on the normal force, the bending force, the lateral force and the twisting force experienced by the probe tip as it scans over the surface. In the AFM *tapping mode*, a sinusoidal optical signal is monitored and the amplitude of this signal represents the level of damping due to the proximity of the surface. In most instruments, the amplitude is kept constant by a feedback mechanism which keeps the probe–surface separation constant. Hence the feedback signal is a measure of the separation or the topographical features on the surface.

In Chapter 1, the spatial resolution limits were discussed in relation to far-field diffraction effects.[25] In 1928, Synge[26] proposed the key feature of what is now known as scanning near-field optical microscopy (SNOM). Consider a tiny aperture illuminated from the back which acts like a sub-wavelength sized light source and which is placed within the optical near-field distance of a transparent sample. (By near-field, the distance between the aperture and sample is less than an optical wavelength.) Then no diffraction limited lens optics are needed for the image formation process and the light collected through the sample carries information of a very small surface volume whose size is largely determined by the probe aperture.

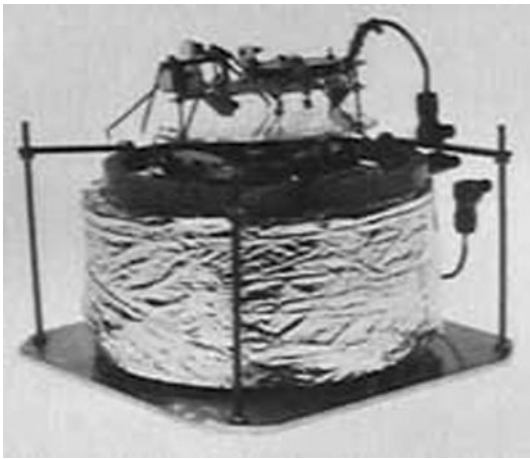


Figure 5.12 Photograph of the original Atomic Force Microscope (AFM) built by Binnig and Rohrer.

By using a tapered optical fibre as either sub-wavelength optical source or as the detector and placing it in the near field, the traditional diffraction limit can be overcome. The first devices operating in the optical regime were built by Pohl *et al.* in the mid-1980s.[27] The near-field optical intensity is not a good feedback signal, so various feedback mechanisms have been employed, including the popular method of monitoring laser light reflected off the probe.

Commercial SPMs tend to scan an area of typically $200 \times 200\,\mu m^2$ and probably height ranges within 10 or $15\,\mu$m, in order to give the user the best spatial resolution. However, one research group has recently described a home-built design – called the 'Milliscope' – with a wide scanning range, which equals the magnification range of an optical microscope.[28] They achieved a large area scan of $1\,mm \times 1\,mm$ with a lateral resolution of 10 nm or less, thus reaching a linear dynamic range of 10^5.

5.3.3 NSOM/SNOM

Near-field scanning optical microscopy (NSOM) – sometimes referred to as scanning near-field optical microscopy (SNOM) – scans a fine-tipped optical probe over the sample surface (see Fig. 5.13). Because the aperture is so small and it is held close to the sample surface, the diffraction effects that limit spatial resolution in conventional optical systems do not restrict NSOM resolution. The NSOM probe is made from single-mode optical fibre that is drawn to an aperture a few nanometres in diameter. NSOM requires a small but constant distance between probe tip and the sample surface. This is often accomplished by monitoring the forces generated between the tip and the sample by shining laser light off the back of the probe. In other systems, a 'tuning fork' approach is used where the probe is oscillated laterally close to the surface.

Amongst the different types of SNOM that have been identified in the literature are the apertureless SNOM (see Kalkbrenner *et al.*,[29] Frohn *et al.*,[30] Zayats and Sandoghdar,[31] Subramaniam,[32] and Hamann *et al.*[33] and the near-field Raman microscope (see Milner and Richards[34]). Typical application areas in materials science for the near-field optical microscope are given in Yamaguchi *et al.*,[35] Inouye and Kawata[36] and Silva *et al.*[37]

The Raman SNOM system built at Leeds is shown schematically in Fig. 5.14 and has been described in a number of papers (e.g. Smith *et al.*,[38] Webster *et al.*[39]). A sub-$\lambda/2$ aperture is used to illuminate the sample and this aperture is a tapered fibre-optic probe coated with 100 nm of aluminium. A $100-200\,nm$ aperture is defined at the apex of the taper. The aperture is held in the near-field using a shear-force feedback technique. The amplitude of the probe's dithering motion is monitored using a simple optical detection scheme. When the probe moves to within 30 nm of the sample, the amplitude of the dither decreases due to the tip-sample interactions and the effect is used by the feedback system to keep the tip-sample separation at around $10-20$nm. As the sample is raster-

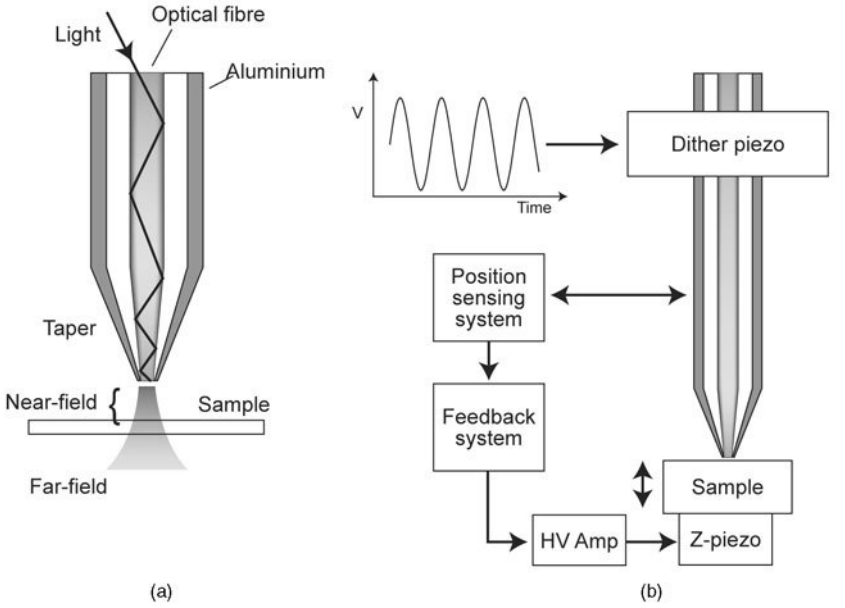


Figure 5.13 Schematic figure showing the basis of near-field microscopy. (a) Light is fed down a tapered optical fibre which has been coated with aluminium. The light in the near-field interacts with the sample and this interaction may be detected either by transmission through the sample or by reflection into the optical fibre. (b) A piezoelectric element may be used to dither the fibre and give surface topological information via a position sensing system.

scanned, the tip follows the contours of the surface so that both topographical images as well as optical images are produced. The sample is scanned in three dimensions under the probe using a piezoelectric positioning stage. The scattered light from the probe-sample region is collected with microscope objectives, either in transmission (via the TCO in Fig. 5.14) or in back-scatter at 60º to the probe axis for opaque samples (via the RCO in Fig. 5.14). The collected light is relayed by a system of lenses into the Renishaw Ramanscope, described earlier in this chapter. Case studies on the typical use of this Raman SNOM for stress measurement in silicon and silicon-based integrated circuits are given in Webster *et al.*,[40] Bonera *et al.*[41] and DeWolf *et al.*[42] Another application area is given in Narita *et al.*[43] Further development of the UV Raman system is being undertaken at Leeds and will be described in Webster *et al.*[44]

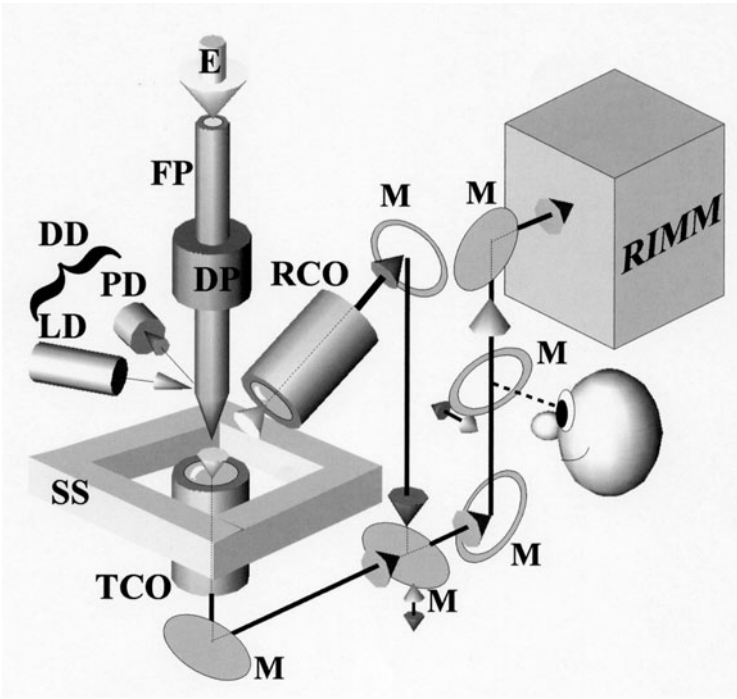


Figure 5.14 Ultraviolet Raman scanning near-field optical microscope (UV RS SNOM) being developed at Leeds by Professor Batchelder's team. The dither detection (DD) of the probe is achieved by the laser diode (LD) and the photodiode (PD). The piezo-scanning stage (SS) moves the sample under the probe and the UV excitation light (E) is partially transmitted through to the transmission collection objective (TCO) and partially reflected into the reflection collection objective (RCO). Various mirrors (M) are moveable so that the operator can observe the probe-sample region through the collection optics. The emergent light exits into the Raman imaging microprobe/microscope (RIMM) for intensity and spectral analysis.

5.4 Miscellaneous optical and EM techniques

5.4.1 Thermography and FT-IR

Although not strictly microscopy, it is interesting to consider the use of infrared radiation in industry for searching for flaws in processed parts. Thermography is the name given to the infrared technique, shown in Fig. 5.15.

This method of non-destructive testing is interesting because it is a large scale and rapid technique, which usually seeks defects within a sample. The principle behind the technique is as follows: all materials radiate energy in the form of EM radiation over a wide wavelength range. The energy radiated per unit area, P, is given by the Stefan-Boltzmann law:

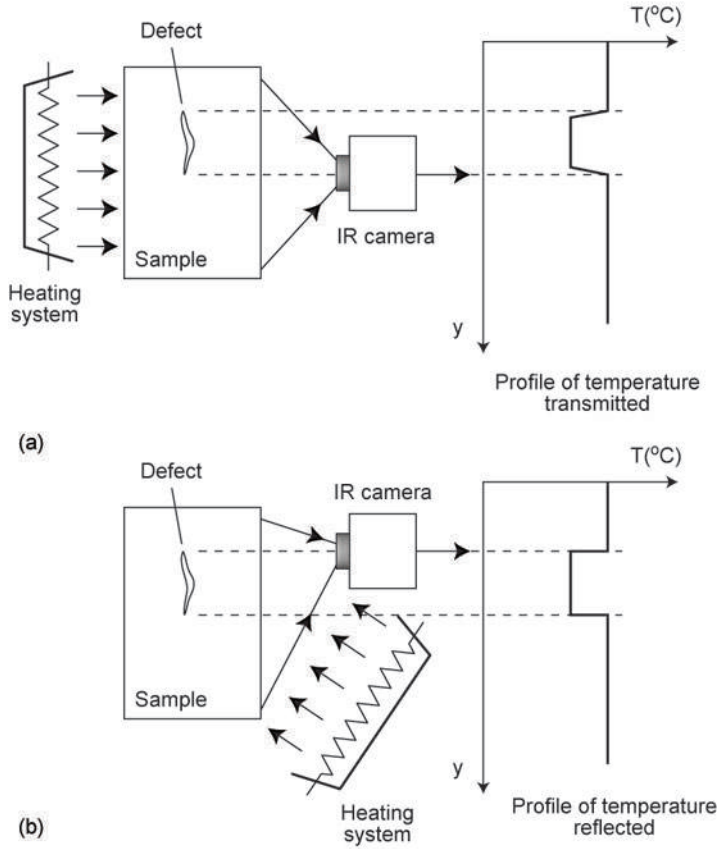


Figure 5.15 The thermography technique for detecting defects within large area samples by means of infrared radiation. The two variations are (a) detection by transmission through the sample and (b) detection by reflection off the sample.

$$P = \varepsilon\sigma T^4 \qquad [5.6]$$

where ε is the emissivity of the surface, σ is the Stefan-Boltzmann constant ($5.67 \times 10^{-8}\,\mathrm{Wm^{-2}K^{-4}}$) and T is the temperature in degrees Kelvin. The emissivity is a fundamental parameter in infrared thermography: it is an intrinsic characteristic of the surface of the material and by definition has a value between 0 and 1. Composite materials, which possess high emissivity, can be tested easily. (Materials with low emissivity should be coated with a high emissivity film to improve the response.) The detector (e.g. indium antimonide) may be placed in contact with a cryogenic container filled with liquid nitrogen (or alternatively a Peltier thermoelectric cooler) to improve the detection sensitivity. The sample may be heated from the opposite side to the detector or heated from the front.

In a slight variation on this theme, IR vibro-thermography is where the sample is subjected to high frequency mechanical vibrations using a piezoelectric vibrator. Internal friction and deformation in the sample causes thermal energy to be radiated. The damaged parts radiate more energy than the rest of the sample giving a detectable signal.

However, arguably the most important infrared spectroscopic technique (other than Raman), that is used in research today, is the Fourier transform infrared (FT-IR) technique (see Schultz[45]). The astronomer, William Herschel discovered the existence of infrared radiation in sunlight and referred to this radiation in a series of four papers, which were presented to the *Philosophical Transactions* in 1800.[46–49] There are basically two methods for recording the spectral characteristics of IR radiation: continuous wavelength selection (traditionally using dispersive gratings) and interferometry. The essential differences between these two types of instrument are illustrated in Fig. 5.16(a) and (b). Nowadays, optical bandpass filters may be employed instead of

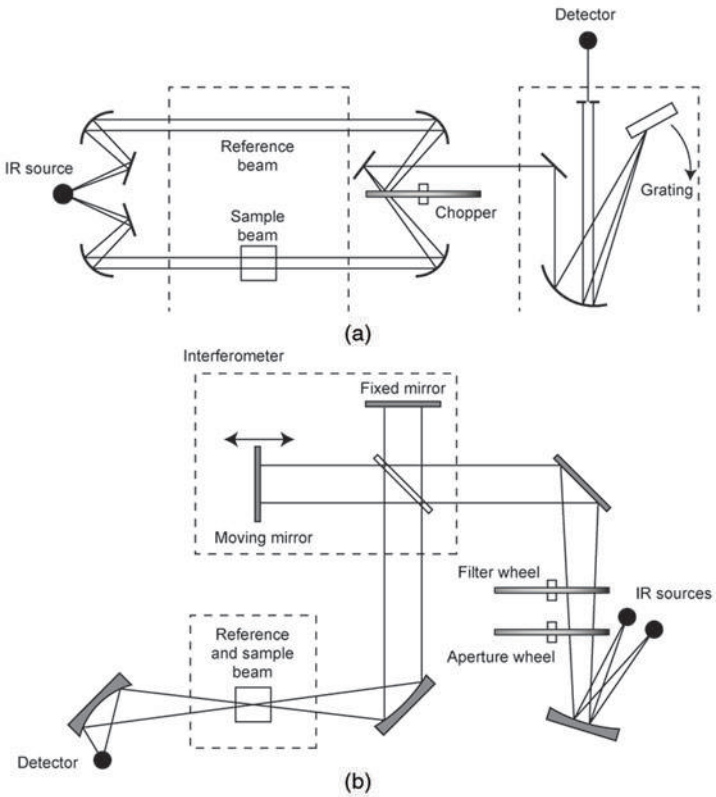


Figure 5.16 Fourier transform infrared detection systems (FT-IR). (a) A dispersive system is shown and, in contrast, (b) an interferometric system.

spectrometer gratings and various tunable filters (liquid crystal or acousto-optic filters) enable IR spectra to be measured much more quickly than previous dispersive gratings. However, Fourier transform infrared spectroscopy (FT-IR spectroscopy) has been steadily gaining in popularity.

Interferometry, coupled with the introduction of computers, has revolutionised this form of IR spectroscopy because the time taken for digitising and fast Fourier transforming (FFT) the raw interferograms to create the IR spectra has fallen rapidly. The interferometer consists of three main parts: a beamsplitter, a permanently-aligned mirror and a moving mirror. As shown in Fig. 5.16(b), the IR is split into two separate beams. The path traversed by one beam, before it recombines with the other (fixed path beam), is made to vary and hence the classic interference phenomenon (see Chapter 1) will take place between these beams. Another interesting example of a Fourier transform infrared system, which was used in a balloon programme in the 1980s for the detection of far infrared radiation, is shown in Fig. 5.17(a). This is a lamellar grating interferometer design (see Birtwell *et al.*[50]) and a typical interferogram created as the moveable plates were stepped through the fixed plates is shown in Fig. 5.17(b). The data points had to be fed into a ground-based computer and the spectral content was deduced by Fourier transformation of the data.

The benefits of interferometry over dispersive technology are threefold. FT-IR is much faster and more sensitive due to the 'multiplex advantage'.[51] It uses all of the IR energy simultaneously and hence achieves better signal-to-noise ratios – this is called the 'throughput advantage'.[52] An internal laser calibrates the interference information and therefore gives very high wavenumber accuracy and reproducibility – referred to as the 'precision advantage'.[53] Also, the instrumental design does not create any spurious, stray infrared radiation, which therefore reduces the inherent noise in the system.

The latest developments in FT-IR spectroscopy relate to the imaging detectors that are now available. Recently, only point detectors (like mercury-cadmium-telluride) were available commercially. Creating images, therefore, could only be accomplished by mechanically (i.e. slowly) scanning the sample within the IR beam. However, there are now focal plane array (FPA) area detectors, which can reduce the time for generating area spectral data by several orders of magnitude. Spectral information that used to take hours or days now takes only minutes. Using focal plane arrays also means that the spatial resolution is given by the pixel size of the FPA, rather than being defined by the diffraction limit through a small aperture within the system.

5.4.2 Fluorescence lifetime imaging (FLIM)

The conventional fluorescence microscope has a number of limitations, for example, absolute quantitative intensity measurements may be degraded by excitation intensity fluctuations, sample heterogeneity, absorption, strong

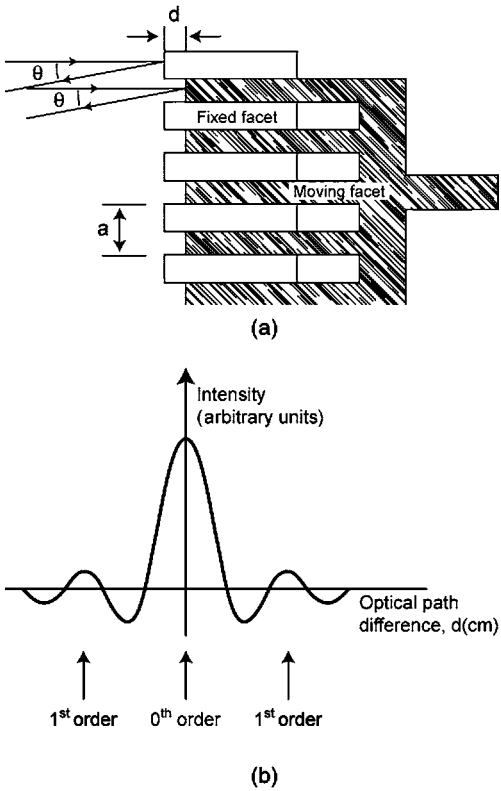


Figure 5.17 (a) The mechanical arrangement of fixed and moveable plates for a balloon-borne lamellar grating interferometer, which was flown to detect far infrared radiation from the galactic centre. (b) A typical interferogram produced by the system as the position, *d*, of the moveable plates varied in relation to the fixed plates is shown. The spectral content of the incident radiation can be deduced by Fourier transforming the interferogram data.

scattering and background fluorescence. These problems could be overcome with a wavelength ratiometric technique when only relative intensity measurements are required.[54] However, even this technique is limited by the small number of suitable fluorescence probes and low signal-to-noise ratios. Fluorescence lifetime measurements, on the other hand, are also derived from relative intensity measurements and may be used to distinguish between fluorophore molecules (due to their different radiative decay rates) and monitor local environmental perturbations that affect the radiative decay rate.

Fluorescence lifetime data may be acquired by frequency- or time-domain techniques. If the sample is excited by a sinusoidally modulated source (typically tens of MHz to GHz) and the fluorescence is sampled using a phase

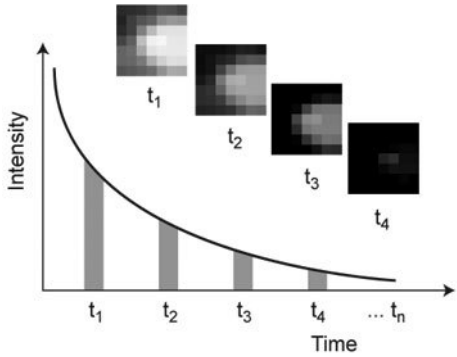


Figure 5.18 The principle behind time-domain fluorescence lifetime imaging microscopy (FLIM) is shown schematically. Time-gated image frames are collected at different time delays t_1, t_2, t_3 and t_4 after the excitation laser pulse. The intensity for each pixel is plotted against time and an exponential curve is fitted. Hence, the fluorescence lifetime may be determined.

sensitive detector, the lifetime can be determined from the phase shift relative to the excitation light source or the modulation depth of the fluorescence signal. If the sample is excited by a short laser pulse, a time-domain measurement is produced, represented schematically by Fig. 5.18, where the resulting decay is sampled at several temporal positions using a time-gated detector. An exponential decay profile is then fitted to the data from each image pixel. The schematic design of a whole-field FLIM system is shown in Fig. 5.19 and is described in detail in Cole *et al.*[55] The key sensing element is labelled the GOI (manufactured by Kentech Instruments Ltd, Didcot, UK), which is a switchable, image intensifier incorporating a microchannel plate. Cole *et al.* have measured time-gate widths of only 90 picoseconds (including jitter). With this narrow temporal resolution, it is necessary to average over many (10–30) frames at each time delay in order to obtain fluorescence images suitable for further analysis. This system has been used in the study of biological samples and a portable, lower cost system has been developed recently using all-solid-state diode pumped laser technology. Cole *et al.* also describe a structured illumination technique, which enables them to achieve 3D optical sectioning. This technique has been applied to fluorescence microscopy too (see Neil *et al.*[56]). Two other interesting recent references related to the FLIM technique are Carlsson and Liljeborg[57] and Straub and Hell.[58]

5.4.3 Dual path programmable array microscope (PAM)

There is an interesting new type of confocal microscope which has only recently been reported.[59] The programmable array microscope (PAM) incorporates a spatial light modulator (SLM) placed in the primary image plane of a wide-field

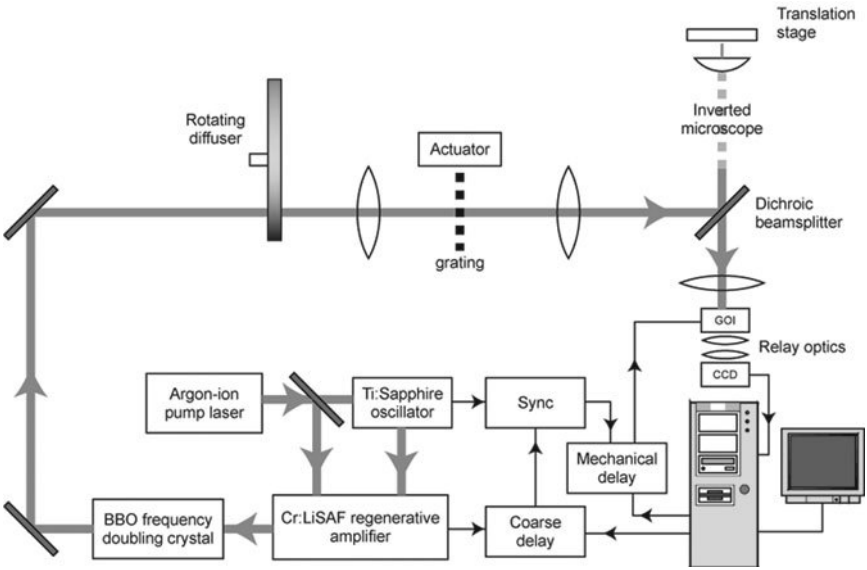


Figure 5.19 An overview of Cole *et al.*'s FLIM system is outlined. An inverted epi-fluorescence microscope is combined with an ultra-fast laser excitation source operating at 415 nm, which produces 1 μJ, 10 picosecond pulses at a repetition rate up to 25 kHz. The fluorescence light is amplified by detector GOI, which is an image intensifier and microchannel plate device, and recorded by the CCD.

microscope, where it is used to define patterns of illumination and/or detection (and see also Chapter 1). The family of PAMs includes optical sectioning systems operating in reflection (see Liang *et al.*,[60]), in fluorescence (see Hanley *et al.*[61]) and spectroscopic imaging systems for fluorescence emission (see Hanley *et al.*[62]).

A new design of PAM, shown in Fig. 5.20, is the dual path PAM which collects both the conjugate image, I_c (formed by light originating from the object plane and returning along the optical path of the illuminating light) and the non-conjugate image, I_{nc}, receiving light from those regions of the SLM that are not used for illuminating the sample. The spatial light modulator element – a digital micro-mirror device (DMD) from Texas Instruments – is effectively a grid of miniature mirrors, which are 16 μm across and separated by 1 μm. The orientation or tilt of each mirror in the grid is controlled electronically. Hence each element of the grid can either reflect light along the conjugate path or the non-conjugate optical path, depending on its tilt. I_c is analogous to a confocal image and I_{nc} is analogous to a conventional image minus I_c. It is claimed that this PAM design provides more time-efficient excitation and also has greater utilisation of the available emission light than the confocal laser scanning microscope.[63]

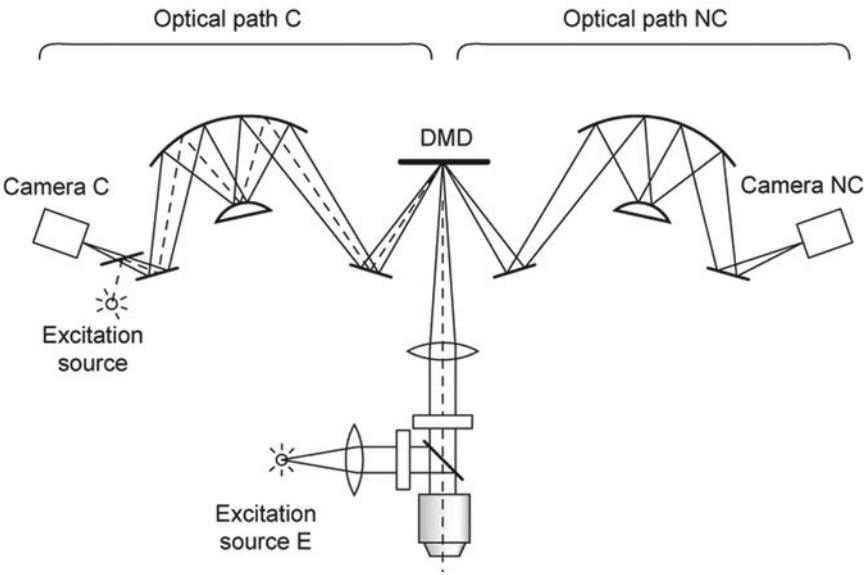


Figure 5.20 Overview of a dual path, programmable array microscope (PAM) is shown. The digital micro-mirror device (DMD) is placed in the primary image plane where it directs light to and from the plane of best focus in the object. The excitation source E may be used to provide epi-illumination. The relay systems are arranged so that focal planes and CCD camera are normal to the DMD.

5.4.4 Optical coherence tomography

Yet another type of confocal technique for 3D reconstruction is optical coherence tomography (OCT), which is a non-destructive and non-contact technique to image microstructure within scattering media (see Dunkers *et al.*[64]). OCT uses light in a similar way to ultrasound imaging using sound and, although it can only penetrate within a few millimetres of a transparent material's surface, it provides higher spatial resolution (typically 5–30 μm) than low frequency ultrasound. OCT is based on low-coherence ranging, i.e. the optical distance to individual back-scattering sites within the sample is determined using an interferometer and a dispersive source. Multiple through-thickness, back reflectance scans are registered whilst transversely scanning either the sample or the probing beam. They are then combined to form a cross-sectional image. As shown schematically in Fig. 5.21, to perform OCT imaging, broad spectrum laser light is transmitted using a single mode fibre and coupled into a 50/50 fibre-optic splitter that illuminates the sample and a linear translating, constant velocity, reference mirror. In effect, the fibre-optic splitter, sample and constant velocity reference mirror constitute a Michelson interferometer. Light back-reflected from each interferometer arm is recombined at the fibre-optic splitter. Interference fringes are registered at the detector only

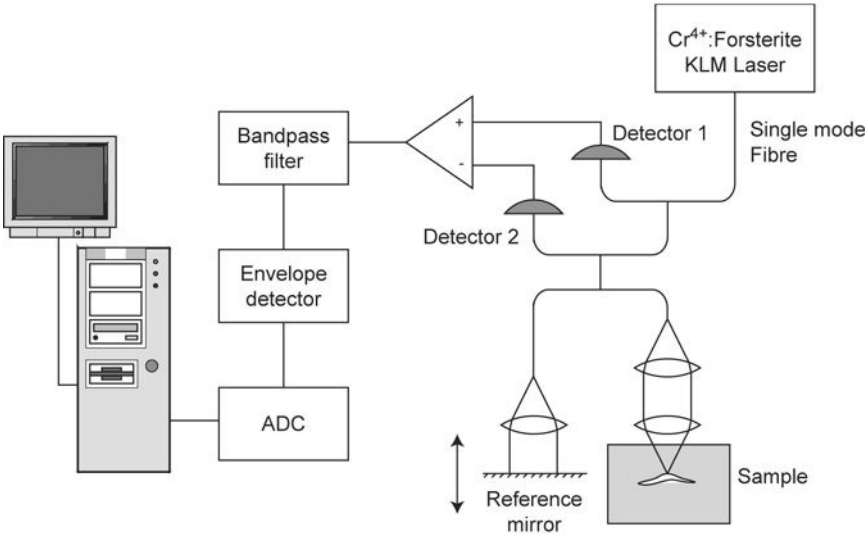


Figure 5.21 In optical coherence tomography (OCT), broad-spectrum laser light (20–200 nm bandwidth) is transmitted using a single mode fibre and coupled into a 50/50 fibre-optic splitter. Both sample and a linearly translating, constant velocity reference mirror are illuminated and form, in effect, a Michelson interferometer. A 25 fs mode-locked, chromium–forsterite, solid state laser may be used and the two detectors allow amplitude noise from the laser to be balanced out.

when the optical path length of the reference arm matches that of the sample arm to within the coherence length (or inverse spectral width) of the source. Therefore, the axial resolution can be as low as 5 μm. Transverse resolution in OCT is determined by the focal spot size of the probing beam, which is usually 10–30 μm. As in confocal laser scanning microscopy, higher numerical aperture optics would provide superior transverse resolution, but at the expense of a diminished depth-of-field. Both OCT and ultrasound suffer from degradation of feature contrast through shadowing of strongly scattering features from above the region of interest in the sample.

Referring to Fig. 5.21, two photodiodes are used for dual-balanced detection (so that amplitude noise from the laser source can be rejected). The interferometric signal was filtered electronically with a band-pass, which was centred on the fringe or heterodyne frequency. The filtered fringe waveform is then demodulated, digitised and stored on a computer. The high dynamic range of this system allows back-reflections as weak as femtowatts of power to be detected. The reference mirror is mounted on a linearly translating galvanometer, which is driven with a triangular waveform. The 260 mm/sec velocity of this mirror results in a fringe modulation frequency of 400 kHz. Transverse scanning was performed using a computer-controlled, motorised

stage to move the sample. Images were created by mapping the logarithm of the signal strength to a greyscale, look-up table. Image acquisition times of one minute were typical. The axial measurement range is determined by the distance the reference mirror moves (4.5 mm) normalised by the refractive index of the sample, n_s, i.e. the axial range is $4.5/n_s$ mm.

5.5 X-ray microscopy/tomography/ microtomography

Clearly, by inspection of the EM spectrum in Chapter 1 and remembering Rayleigh's criterion, a dramatic improvement in spatial resolution should result if an X-ray microscope could be devised (because the wavelength of soft X-rays is in the 2–5 nanometre range). Note that X-ray transmission through materials is limited by the density, size and atomic number of the target material. In Fig. 5.22 are shown typical attenuation factors for the X-ray EM regime. Sources of X-rays are many and varied: from massive particle accelerators for some X-ray microscopy applications through to small portable sources for table-top

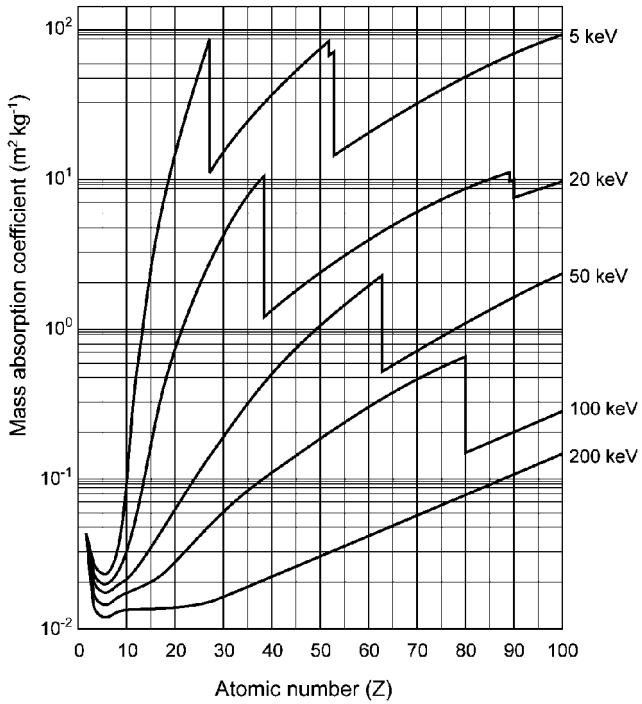


Figure 5.22 Some typical curves of the mass absorption coefficients for different atomic number target materials, at various incident X-ray energies.

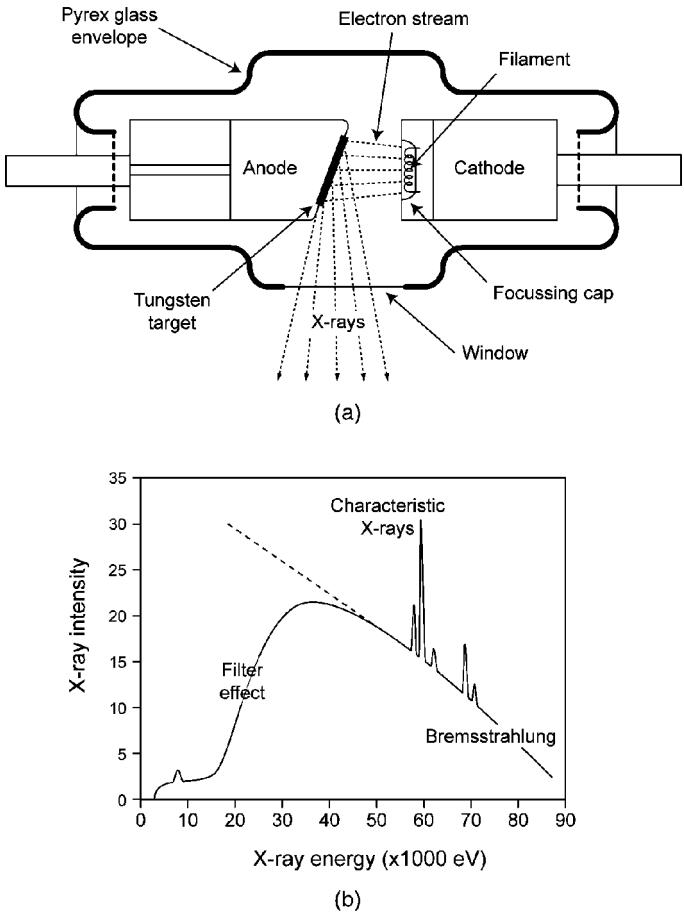


Figure 5.23 (a) A typical X-ray generation method is shown where high energy electrons from a heated filament are attracted to and strike a tungsten target. The target must be cooled to prevent overheating and the *X*-rays escape through the thin window in the pyrex envelope. (b) An illustration of a typical X-ray spectrum where not only characteristic emission lines due to the target material are generated, but also broad band, 'bremsstrahlung' or 'braking radiation'. The characteristics of the window will determine exactly where the filtering effect takes hold in the spectrum.

installations of X-ray micro-tomography. The mechanism for X-ray generation is often electron beam collisions with a high-Z target. Hence, there are many X-ray detection systems in operation on electron microscopes, e.g. XPS, ESCA (and these will be discussed briefly in Chapter 6). Figure 5.23 shows an example of a typical X-ray tube.

5.5.1 Soft X-ray microscopes

Compared to electrons, soft X-rays penetrate much thicker samples, up to about 10 μm, or the full thickness of a sectioned cell in biology. Soft X-rays interact with matter to produce high contrast biological images at atmospheric pressure based on intrinsic components (rather than stained components). By appropriate selection of illuminating wavelengths, this contrast can be both element-specific (see Kenney *et al.*[65]) and chemical-specific (see Zhang *et al.*[66]). The only problem is that these X-rays are very penetrating and safety issues regarding their production and containment are non-trivial. Indeed, the only beam source of such X-rays is at special particle accelerator installations like the Lawrence Berkeley National Laboratory in the USA (Meyer-Ilse *et al.*,[67,68]) and Gottingen in Germany (Niemann *et al.*[69]). When elementary particles are constrained in circular orbits within the storage rings of the particle accelerators, they emit a characteristic radiation – called synchrotron radiation – and are a reliable source of these soft X-rays. Note also that interesting claims are being made for a 3D crystal X-ray microscope developed at the Oak Ridge National Laboratory in Tennessee, USA.[70] This instrument can probe inside a polycrystalline material in situ, allowing researchers to see the effects of applied stresses and strains on individual grains. One of the many innovations that form a part of the instrument design is the fabrication of a Kirkpatrick-Baez X-ray mirror, which is capable of focussing the high energy X-rays (from a beam line like the advanced photon source at Argonne National Laboratory) to a sub-micron spot size. However, in smaller scale size X-ray instruments for tabletop work, instead of special diffracting or reflecting optics, zone plates are generally used to focus the X-rays.

So far, when discussing diffraction, it has been assumed that the diffracting aperture was illuminated by parallel light and that an objective lens was an intrinsic part of the experimental arrangement (so that the results would be relevant to optical microscopy). However, when the diffraction of light through an aperture is created by a finite distance source and the light diffracted is viewed directly on a screen, which is a finite distance from the aperture, Fresnel diffraction theory (rather than Fraunhofer diffraction) must be invoked. The contributions of all secondary wavelets within a rectangular aperture may be evaluated at a point *P*, which is close to the aperture, by summing on a complex plane. The resulting figure is a Cornu Spiral, as shown in Fig. 5.24(a). Consider the situation in Fig. 5.24(b) and the effect of different parts of the wavefront on the image at *P* on the screen. Clearly, using Huygens secondary wavelets concept, the phase difference, ϕ, between the two light paths shown is given by:

$$\phi = \frac{2\pi}{\lambda}(r_p - (PS - r_s)) \approx \frac{\pi(PS)}{r_s(PS - r_s)\lambda}a^2 \qquad [5.7]$$

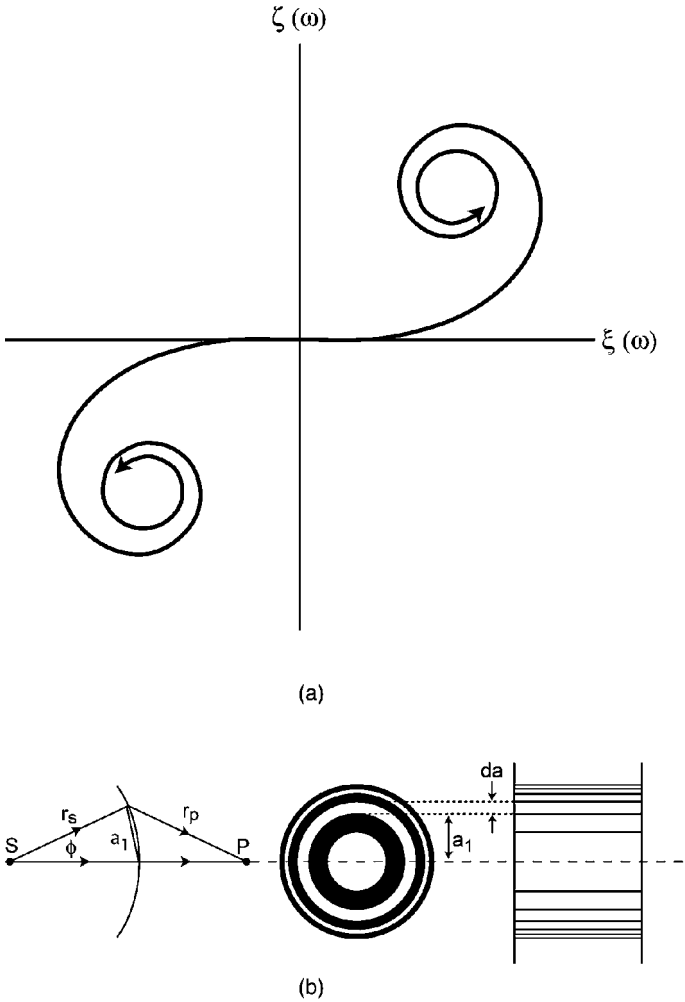


Figure 5.24 The resultant intensity at a point, P, near to a rectangular aperture is more complex to calculate than Fraunhofer diffraction at points far away from the aperture and Fresnel integrals must be used. (a) The Cornu Spiral is formed from the amplitude and phase of different secondary wavelets within the rectangular aperture, where the axes of the complex plane representation are $\zeta(\omega) = \int_0^{\omega} \sin\left(\frac{\pi\omega'^2}{2}\right) d\omega'$ and $\xi(\omega) = \int_0^{\omega} \cos\left(\frac{\pi\omega'^2}{2}\right) d\omega'$ and it enables intensities to be determined. (b) The nature of the varying width (but constant area, *da*) zones on a zone plate are shown, where, by blocking off specific zones which are in phase with each other, the zone plate is made to behave like a lens with focal length, $f = a_1^2/\lambda$ where a_1 is the radius of the first zone.

Therefore the phase difference, $\delta\phi$, across an elementary annulus (or zone) of radius a and width δa will be obtained by differentiating the following expression:

$$\delta\phi = \frac{2\pi(PS)}{r_s(PS - r_s)\lambda} a\delta a \qquad [5.8]$$

For these annular zonal elements, the areas are made the same, so that each has area $2\pi a\delta a$ and hence the contribution to the amplitude from each element should be the same.

A zone plate is formed when the light from every half-period zone is blocked off, i.e. zones are marked off on a transparent surface with radii proportional to the square roots of whole numbers and alternate zones are blackened (see Fig. 5.24(b)). This drawing is then photo-reduced and when EM radiation is passed through the reduced zone plate, an intense point image is produced similar to the image produced by a lens. It can be shown that the effective focal length, f_{zp}, of the zone plate is given by

$$f_{zp} = \frac{a_1^2}{\lambda} \qquad [5.9]$$

where a_1 is the radius of the first zone. The zone plate is widely used in X-ray microscopy for focussing the rays (without the attenuation associated with refracting lens material).

Although most transmission X-ray microscopes rely on high brightness synchrotron radiation sources in order to achieve short exposure times, Berglund *et al.* have recently reported a compact tabletop soft X-ray system with sub-100 nm resolution and operating in the water window wavelength range ($\Delta\lambda = 2.3 - 4.4$ nm).[71] A schematic view is shown of their novel equipment in Fig. 5.25, which uses a droplet target system.

In a typical arrangement, laser pulses from a 10 Hz 100 picosecond frequency-doubled Nd:YAG laser are focussed onto 15 μm diameter ethanol droplets, which are generated by a piezoelectrically-vibrated glass capillary nozzle. The laser-produced plasma acts as a high-brightness X-ray source, emitting over 4π steradians. Therefore a high collection efficiency condenser optical design is necessary to obtain sufficient photon density in the object plane for reasonable exposure times. In their system, a spherical multilayer condenser mirror, matched to the numerical aperture of the diffractive zone plate, directs the X-rays through the sample. The mirror had two hundred layers of W/B_4C sputtered onto its surface, each with a 2D spacing corresponding to first-order, normal-incidence reflectivity at 3.37 nm of 0.5% and these thin films created a narrow bandwidth around the chosen line of only 0.04 nm. In this way, image degradation due to other emission lines from the plasma was minimised.

Note the central stop placed above the source to prevent the direct light from the plasma (which would saturate the CCD detector) and which also eliminates

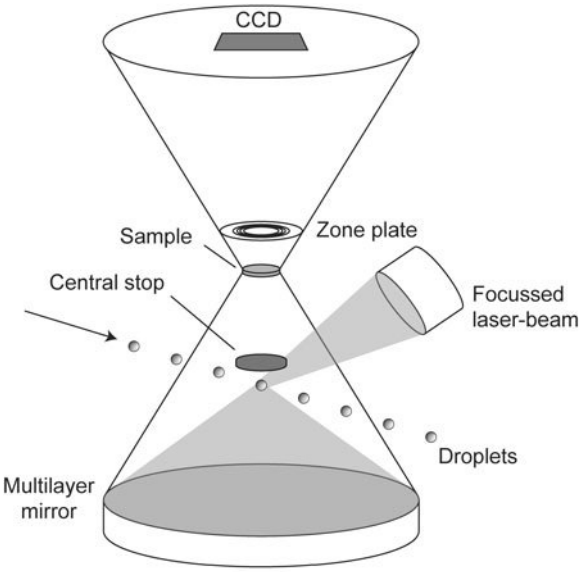


Figure 5.25 Berglund's apparatus for soft X-ray research consists of a regenerative, droplet-target, laser plasma X-ray source; a normal incidence, multilayer, condenser mirror; a sample and a zone plate for high resolution imaging on a back-illuminated, soft X-ray sensitive CCD detector.

the zero order radiation from the nickel zone plate. The CCD is a cooled, thinned, back-illuminated, 1024×1024 pixel array. The overall magnification of their system depends on the distance between the zone plate and the CCD detector. They claim to be able to image features less than 60 nanometres with this system. However, for best results, the sample must be of limited thickness (less than a few microns) because, at the chosen wavelength of 3.37 μm, carbon-based substances exhibit significant absorption.

5.5.2 X-ray microradiography

The development of the confocal optical microscopes has meant that limited penetration (tens of microns) into many semi-transparent materials is now attainable – but these depths are limited by the working distance of the objective lens, even for perfectly transparent materials. X-rays are the natural probe for maximum penetration of material structure and the historical development of methods to determine fibre orientations in composite materials illustrates the importance of X-rays. One of the earliest papers on fibre orientation distributions (Darlington *et al.*[72]) used X-ray contact microradiography which was capable of imaging all fibres within a 100 μm thick slice of the sample. Typical regions studied were 600 μm $\times$ 600 μm $\times$ 100 μm in X, Y and Z. Preferred fibre directions in the XY plane were

then obtained by manual analysis of the overlapping glass fibres in polypropylene and polyamide 66. In high packing fraction fibre-reinforced composites it is difficult to separate the fibres and impossible to obtain true 3D characterisation by contact microradiography. However, more recent investigations have been conducted into 3D, X-ray scanning tomography (see, for example, Geier[73]). Medical scanners operate with spatial resolutions of the order of 200 μm, but enhancements can be made to reduce the voxel resolutions down to 10 to 20 μm, as discussed in Maisl *et al.*[74] Nowadays, compact, tabletop X-ray microtomography systems are available commercially and they show considerable promise for 3D reconstruction of many diverse materials, as discussed in the case study below.

5.6 Case study: X-ray microtomography of fibrous structures

5.6.1 Summary

In Chapters 3 and 4 sample volume reconstructions were approached using a serial sectioning schema (both physically and optically). In this case study sample volumes are reconstructed from a number of X-ray 'shadow' images using computed tomography techniques. The use of X-rays dramatically increases the penetration depth resulting in the reconstruction of much larger sample volumes (compared to visible light microscopy techniques).

5.6.2 Introduction

X-rays were discovered in 1895 by Wilhelm Roentgen, while working in his darkened laboratory. His experiments involved the study of cathode rays produced by discharging electrical current through an evacuated tube. During one such routine experiment he noticed a surprising glow on a barium platinocyanide screen some distance from his apparatus. He was probably even more surprised to discover that when he placed his hand in front of this screen, he could clearly make out the bones in his hand! The identity of the rays that caused this astonishing effect remained a mystery for many years, hence they were given the somewhat ambiguous label of 'X'. Roentgen's early radiographs, as shown in Fig. 5.26, immediately caught the public's imagination and led to Roentgen earning the first Nobel Prize in Physics in 1901.

Figure 5.23 showed a diagram of a typical X-ray tube, which is still a common means of creating X-rays within a laboratory. Electrons are emitted at the cathode and are accelerated rapidly to high speeds by the applied voltage before striking the anode target. When the target material (typically tungsten) is struck by high energy electrons, X-rays are emitted over a broad spectrum of wavelengths.

When X-rays pass through matter they produce the same phenomena, e.g. diffraction, refraction, fluorescence, as does visible light, which were discussed



Figure 5.26 The famous radiograph made by Roentgen on 22 December 1895, which is traditionally known as 'the first X-ray picture'. It is a radiograph of Mrs Roentgen's hand.

fully in Chapter 1. However, the design of an X-ray imaging system is difficult because of their characteristically short wavelengths. Firstly, X-rays are largely absorbed rather than reflected, therefore focussing with mirrors is not a viable solution. Secondly, the index of refraction, n, is ~ 1 at these wavelengths for most materials. Hence, any material sufficiently thin to transmit X-rays would

possess a very long focal length (because the focal length, f, is proportional to $1/(n-1)$, see Chapter 1, equation [1.51]), making the construction of magnifying optics impracticable. However, the Fresnel zone plates, consisting of concentric rings (as discussed in Section 5.5.1) can be used to focus X-rays by diffraction rather than refraction. X-ray microscopes, which exploit this effect, have been developed recently.[75]

It is clear from its use for medical radiographs that X-rays are able to penetrate much further into matter than visible light. Therefore, this property could be exploited to non-destructively image large volumes of material without the inherent problems faced by serial sectioning (see Chapter 3) or depth of penetration of visible light (see Chapter 4). However, without a means of focussing the illuminating X-rays, an alternative approach is required to resolve features within the sample.

5.6.3 X-ray computed tomography

When X-rays pass through a sample, they are absorbed according to spatial variations in composition and density providing the necessary contrast seen within radiographs (or shadow images). Figure 5.27 illustrates the creation of a shadow image formed when a parallel X-ray beam passes through the sample. It is not possible to determine from a single shadow image where, within the sample, this attenuation has occurred. However, if the measured attenuation is distributed evenly along the line of projection (i.e. parallel to the X-rays), the

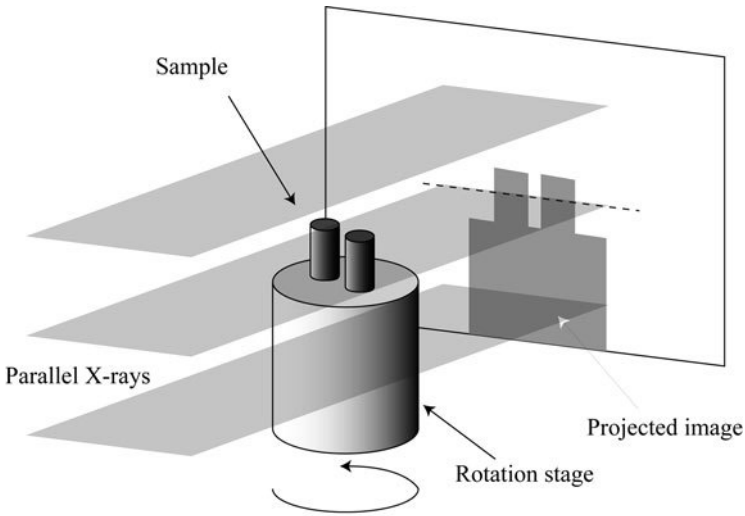


Figure 5.27 During X-ray tomography, parallel X-rays pass through the sample resulting in the production of a shadow image. The dotted line indicates the section reconstructed by back projection in Fig. 5.28.

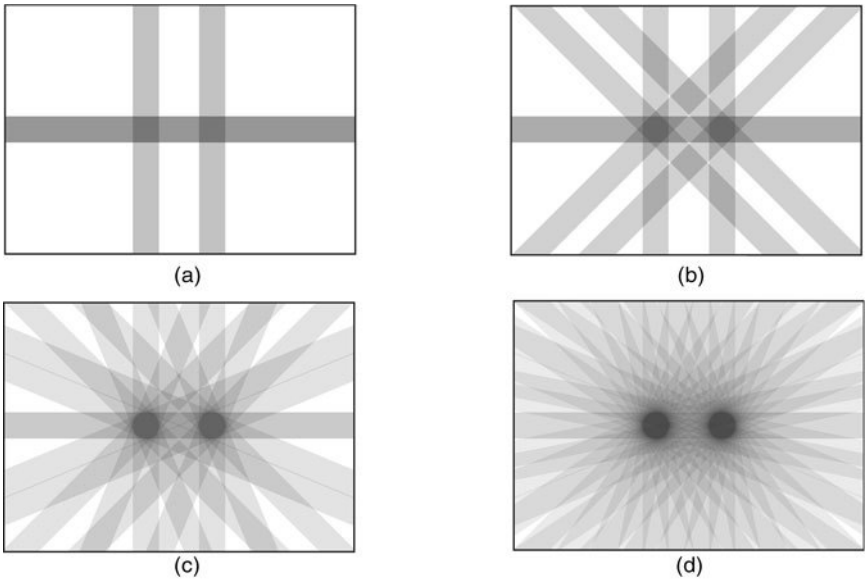


Figure 5.28 The reconstruction of a single sample cross-section by back projection. The observed line intensity from each X-ray shadow image is projected across the reconstructed image at the respective sample orientation. Reconstructions from 2, 4, 8 and 16 projections are illustrated in (a)–(d) respectively.

superposition of attenuation values, as more projections are combined, should correspond to the features present within the sample. During the tomographic reconstruction process, shadow images of the sample are collected as it is rotated by discrete angular increments. Individual line intensity profiles are superposed in order to reconstruct sections through the sample volume.

The reconstruction procedure based on this 'back projection' principle is illustrated in Fig. 5.28. The observed line intensity from a single X-ray shadow image is projected across the reconstructed image at the respective sample orientation. As more shadow images are collected the resulting projections are superposed and the reconstructed cross-section image improves. In this example, the sample consists of two cylindrical objects separated by a small distance. The structure of the sample is clearly visible in the reconstructed cross-section after the back projection of 16 images. However, there is considerable blurring present at the objects' boundaries, which the collection and back projection of further images will reduce, but not entirely eliminate. Various techniques exist for altering the line profiles in order to enhance the edges of features, further reducing this effect.[76] Also, alternatives have been proposed to the back projection techniques which, although more computationally intensive, provide higher quality reconstructions.[77]

The earliest computed tomography (CT) scanners were developed and used extensively in medical applications in the 1970s. However, with the computer processing power available at the time, the reconstruction of cross-section images took many hours. With current CT scanners, computer power is no longer a prohibitive factor. This has allowed the collection and back projection of a far greater number of shadow images (and their line intensities respectively), resulting in much clearer tomographic sections.

Modern medical CT scanners have spatial resolutions of the order of millimetres. However, the recent development of higher-powered, desktop, X-ray tubes, has led to considerable improvements in spatial resolution. High resolution computed microtomography (μCT) systems are now available with manufacturers claiming resolutions of down to 2 μm. Desktop μCT scanners utilise microfocus X-ray tubes which produce a cone of light, allowing variable magnification as the sample is moved towards or away from the X-ray source.[78] The X-ray image of the sample is focussed onto a scintillating material, which, as it fluoresces in the optical part of the EM spectrum, can then be imaged by a standard CCD camera.

5.6.4 Reconstructed image quality – artefacts and resolution

The manner in which a sample is reconstructed with the CT technique is very different to that of conventional or confocal microscopy. Hence, the factors which dictate the reconstructed image quality are also very different. An understanding of the artefacts that may occur in the CT reconstruction is important if they are to be recognised and possibly eliminated. A few of the more common CT artefacts are highlighted below. A more thorough discussion of CT artefacts and algorithms for their correction can be found elsewhere.[79]

Figure 5.29(a) shows a reconstructed cross-section image of a low density felt sample. During the shadow image acquisition, the sample changed shape slightly as it 'relaxed', resulting in a blurring of the reconstructed images. Sample stability is very important during tomographic imaging; fortunately this is not a problem for most samples as they will be quite stable over the time-scales required (about one hour).

Figure 5.29(b) shows a section through a polymer composite sample where rings of varying intensity are seen emanating from the rotation axis of the sample. This artefact is due to slight variations in collection efficiency between neighbouring pixels on the CCD camera which collects the shadow image. The effect can be suppressed if the sample is randomly translated by a known amount between each image exposure. Therefore different pixels on the CCD camera will image the same part within the sample.

Figure 5.29(c) shows the same sample as in Fig. 5.29(a). However, in this cross-section image, the felt has been compressed between two metal plates held together by three steel pins. It can be seen that the presence of the relatively high

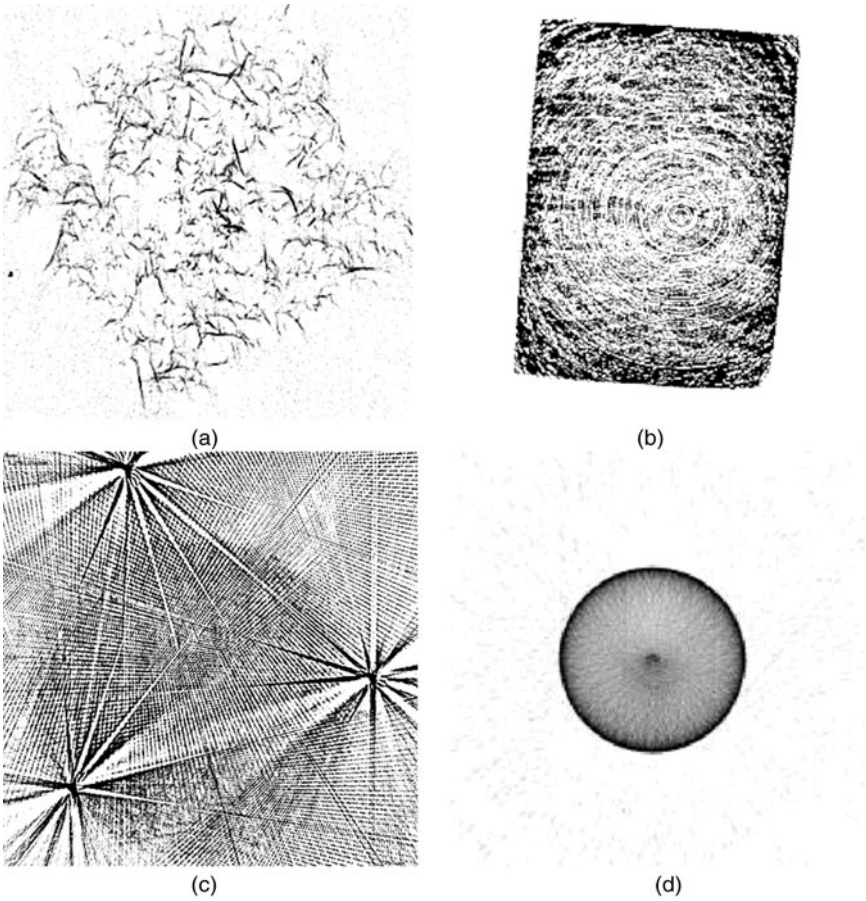


Figure 5.29 Common artefacts are shown that may be present in CT images. (a) The sample relaxed and changed shape during the exposure, giving blurring. (b) Ring artefacts may be caused by variations in CCD pixel collection efficiencies. (c) Star-like artefacts may be produced when high density objects are within the field of view. (d) Edge darkening artefacts may be due to an effect called 'beam-hardening'.

density pins decimates the image of the lower density felt. Because of its higher atomic number, the metal attenuates X-rays considerably with very few passing through the pins into the surrounding sample. Furthermore, if the image of the pins saturates the shadow image, any information on the attenuation due to the felt sample is entirely lost.

Figure 5.29(d) shows a cross-section image of a brass cylinder where the edges are much darker due to an effect called 'beam-hardening'. X-ray tubes like the one illustrated in Fig. 5.23 produce a broad spectrum of light (i.e. they are polychromatic). The X-ray attenuation of an object is wavelength dependent

and therefore, as an X-ray beam traverses an object, the higher energy portion of the X-ray spectrum survives whereas the lower energy photons suffer higher attenuation. This effect can be reduced if the X-ray illumination is filtered by a monochromator. However, the overall intensity will be reduced significantly, resulting in the need for much longer exposure times.

In Chapters 1 to 4 we saw that the spatial resolution of both a conventional and a confocal light microscope is dictated by the focussed spot of the objective lens, i.e. the Airy pattern. Therefore, the resolution of an image is dependent on the chosen objective lens and also the path the light takes through the specimen (which may aberrate the focal spot). In μCT the resolution of the shadow images can be measured using conventional methods, i.e. by imaging gratings or imaging sub-resolution objects. However, the quality of the cross-section images depends on a number of other factors including the number of rotation steps and the accuracy of the back projection algorithm; hence spatial resolution within CT images is not so easily defined.

5.6.5 Fibrous structures in textiles and composites

Fibrous structures are an important component of biological, organic and inorganic materials, which are abundant in our everyday lives. Examples include materials which are inherently fibrous such as textiles and those where the fibrous structure is present within a second phase such as in fibre-reinforced composite materials. The characteristics of the (often-microscopic) fibrous structure itself, e.g. length, diameter, orientation and curvature distributions, play a primary role in dictating the macroscopic properties of the material. Therefore, in order to gain an understanding of these bulk properties, (e.g. elasticity, strength, permeability, thermal conductivity) the microscopic fibrous structure must be probed and characterised. In this case study, the fibrous structure of a number of different composite and textile samples will be studied using μCT.

Textiles are fibrous materials which have found numerous applications in both domestic and industrial environments. The arrangement of the fibres within the textile material can be manufactured in a variety of forms from the very orderly, woven or stitched materials to the less ordered, non-woven, felt-like materials. The standard method for characterising textiles is to acquire a surface image using a light or electron microscope and these images are then analysed using standard image processing methods, for example, skeletonisation (thinning) and chord tracking,[80, 81] or the Hough transform,[82] as in the following example.

Figure 5.30(a) shows an image of a non-woven textile sample, which was acquired using a flat-bed scanner. The Hough transform has been applied to this image (after edge detection using a Laplace filter) and peaks in the parameter space above a predetermined threshold have been identified. The number of

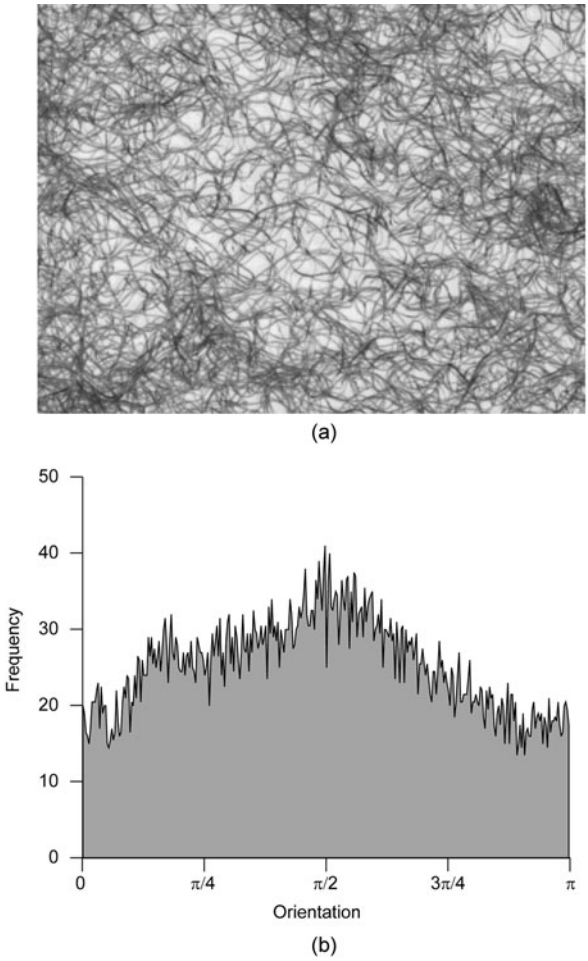


Figure 5.30 (a) An image frame of a non-woven textile sample taken with a flat-bed scanner and (b) the frequency distribution of fibre segment orientations, which have been derived using a Hough transform (see Chapter 2).

peaks within discrete orientation ranges is counted, giving the orientation frequency distribution illustrated in Fig. 5.30(b). The sample illustrated has been elongated manually to twice its original length in the X direction. This distortion has resulted in an anisotropic fibre orientation distribution as illustrated.

The reconstruction of the fibrous microstructure of composite materials has been discussed in some detail within the case studies of Chapters 3 and 4. To recap briefly, fibre characterisation is often limited to the study of physical cross-sections and serial sectioning (Section 3.3) or limited-depth volume reconstructions (Section 4.5.2). A similar problem is faced in both textiles and composites research, where the majority of materials characterisation is

restricted to 2D sections. If the material is sufficiently thin, two-dimensional characterisation may be sufficient. However, if the material has a significant thickness or the parameter of interest cannot be measured from a single section, then time-consuming serial sectioning is usually the only alternative available.

5.6.6 Microtomography of fibrous structures

Figure 5.31 shows three different samples which have been analysed using the Skyscan 1072 μCT. For each sample, the X-ray shadow image is illustrated together with one single, reconstructed, cross-section image. The time taken to acquire and reconstruct each of these volume datasets was approximately 2–4 hours.

Figure 5.31(a) shows a two-folded acrylic yarn, i.e. it is composed of two single threads twisted together. Each single thread has an anti-clockwise twist, which is counter to the clockwise 'folding' twist. The threads were spun out of cut acrylic fibres which have lengths of a few centimetres. It is the friction between neighbouring fibres and the twist of the threads which holds the yarn together. There is considerable interest in the path that individual fibres take through the yarn. Fibres are found to 'migrate' from the centre[83] to the boundaries. These fibres are often modelled as helices.[84] In the cross-section image, shown in Fig. 5.31(b), the position of the individual fibres is clearly visible and the migration of fibres can be observed over a series of sections.

Figure 5.31(c) shows a fibre-reinforced composite containing a 13% volume fraction of short glass fibres (mean diameter, $D = 13\,\mu$m) in a polypropylene matrix, which has been cut to give a cross-section area of approximately $2.3\,\text{mm} \times 2.3\,\text{mm}$. The sample was fabricated by injection moulding, where a mixture of molten polymer and fibres is forced rapidly into a mould at high pressure. There is considerable interest in the effect of the moulding conditions on the resulting fibre orientation and length distributions.[85] Current research interest in computer simulations of injection moulding is starting to produce predictions of the 3D fibrous microstructure. However, measurements of this structure within real samples are still restricted to 2D sections (see Chapter 3). In the reconstructed X-ray μCT cross-section image, which is shown in Fig. 5.31(d), individual fibres are clearly resolved. These data allow for the quantitative characterisation of the spatial or orientation state of the fibres from numerous non-destructive sections.

Figure 5.31(e) shows a polyarylamide non-woven felt sample, with a mean fibre diameter of 42 μm, that has been needle-punched to encourage fibre entanglement.[86] The fibres within thin non-woven webs tend to be oriented within the plane of the web. The process of needle punching, as the name implies, involves rapidly penetrating the felt with an array of needles in order to encourage through-thickness orientation. In order to assess the effect of this processing technique with a light microscope, samples must be set in a resin,

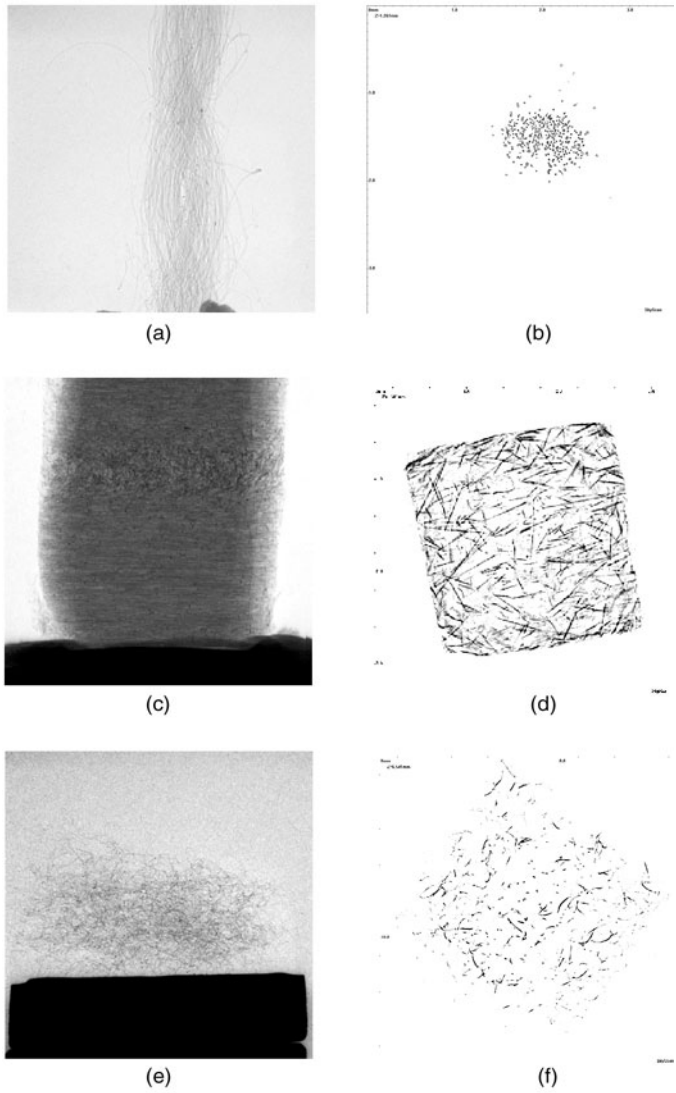


Figure 5.31 Three different materials in the process of reconstruction: (a) and (b) show a textile yarn, (c) and (d) show a glass fibre-reinforced polymer composite and (e) and (f) show a piece of felt. The scale size of the cross-sections (a), (c) and (e) cover an area of approximately a few mms $\times$ few mms and similarly for the area cross-section shown in (b), (d) and (f).

then polished and sectioned as with composite samples. This is clearly a lengthy process which is prone to process-induced bias. A single reconstructed cross-section through this sample is shown in Fig. 5.31(f), demonstrating how μCT can be used to acquire non-destructive sections through non-woven samples.

5.6.7 Automated fibre tracing

The maximum number of cross-sections which can be reconstructed for a particular sample is dependent on the pixel dimensions of the shadow images. With the Skyscan 1072 microtomograph, the shadow images are 1024×1024 and it is possible to reconstruct a cross-section for each line of pixels in the Y direction, giving a maximum of 1024. Each of the reconstructed cross-sections is itself a 1024×1024 image, therefore the space occupied by the reconstructed volume is $1024^3 \times 8$ bits, i.e. one gigabyte. In Fig. 5.32 only a single cross-section has been illustrated. However, for each sample, some 200–500 other cross-sections were also reconstructed.

With the number of images typically acquired using μCT, manual processing (for example, to identify the fibres) would be undesirable. However, many of the 2D image processing techniques do not translate straightforwardly into three dimensions. For example, most skeletonisation routines rely on preserving the connectivity of a region. In 2D, each pixel is connected to 8 neighbours giving 2^8 different combinations, whereas in 3D, each pixel has 26 neighbours, which increases the number of combinations to 2^{26}. Clearly this situation introduces a colossal increase in computation time. Therefore, care must be taken in selecting suitable image processing algorithms and their operations must be optimised.

A novel technique for locating fibres within 3D μCT datasets has been developed. A brief summary of the steps involved is given below (and these steps are described in more detail in Eberhardt and Clarke[87]):

1. *Determination of local orientation* – The most probable orientation (θ, ϕ) is determined at each point (x, y, z) within the image. This is achieved by calculating the mean voxel intensity across a vector (of l voxels in length) at various orientations, as illustrated in Fig. 5.32(a). If a fibre is located at the point (x, y, z), then the sharp peak in voxel intensity can be used to determine the fibre orientation at this point, as illustrated in Fig. 5.32(b). The orientation at each point (whose intensity is above a predetermined background level) within a single cross-section is shown in Fig. 5.33(a).
2. *Collapsing the vectors* – It can be seen from Fig. 5.33(a) that the vectors which lie along a fibre are parallel, whereas those which are located either side of the fibre are perpendicular. This property can be used to eliminate the majority of the vectors, which do not lie on a fibre, as illustrated in Fig. 5.33(b).
3. *Connecting the vectors* – The neighbours of each vector are considered in turn and connections are made between them if they are aligned. As more vectors are connected, the paths of individual fibres are determined, as illustrated in Fig. 5.33(c).
4. *Smoothing* – The polynomial fitting scheme, which was applied to wavy fibres in Chapter 4, was applied to these fibres giving a continuous description of their centre location, as illustrated in Fig. 5.33(d).

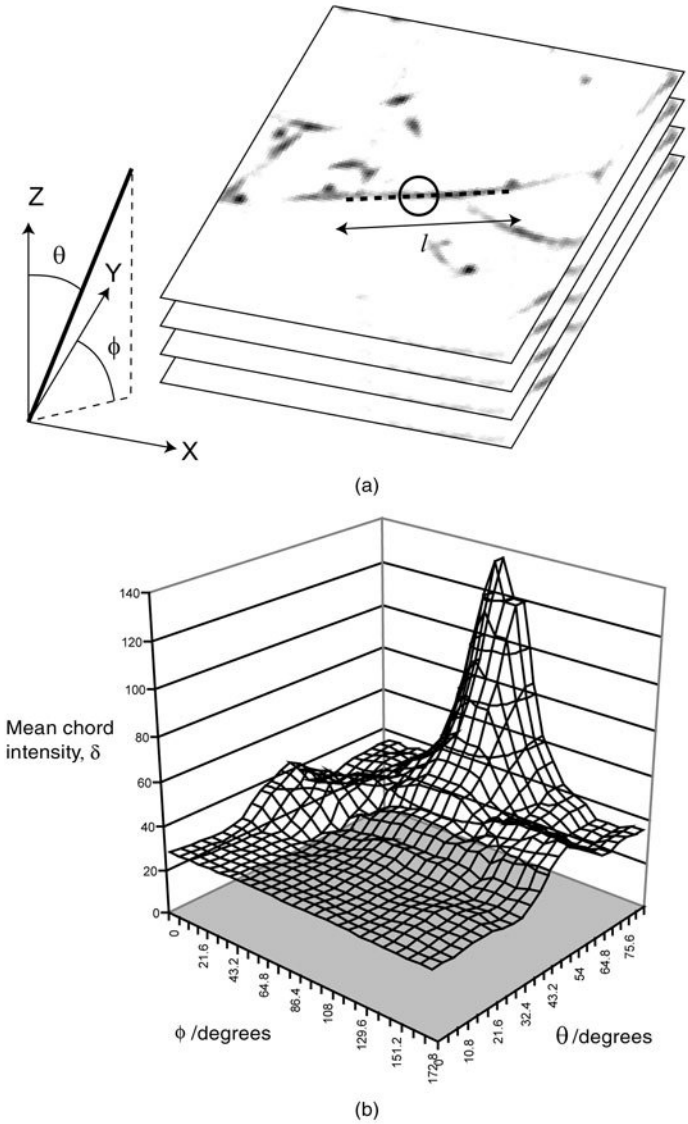


Figure 5.32 (a) An illustration of the parameters used in the fibre determination routine and (b) voxel intensity as a function of the fibre segment orientation values yields a best-fit (θ, ϕ) for that segment.

The results of fibre-tracing the yarn, composite and non-woven datasets (shown in Fig. 5.31) are illustrated in Fig. 5.34. Within the composite dataset the total reconstructed fibre length, found by summing the length of all the fibre segments, is 1.6 metres. This fibre length gives a total fibre volume fraction of 5%, which

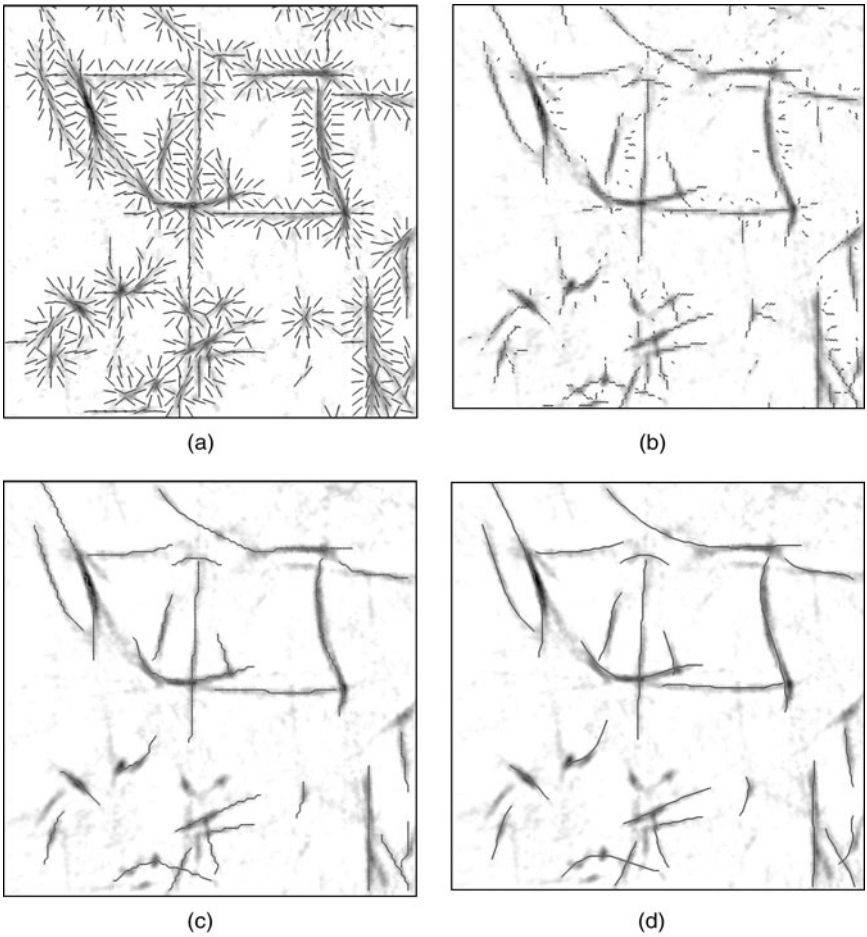


Figure 5.33 The succession of steps for joining the most likely fibre segments and ignoring the spurious fibre segments is shown. (a) and (b) show the effect of collapsing the vectors, (c) illustrates the connecting of the most likely vectors and (d) shows the vectors after a polynomial smoothing has been applied.

compares to the actual volume fraction of 13%. Therefore, the fibrous structure illustrated in Fig. 5.34(b) shows slightly less than half of the material in the sample! Similar methods can be applied to determine the amount of missing data in the other two datasets. There are thought to be two reasons for these missing data. Firstly, it is dependent on the quality of the cross-section images (which is dependent on the resolution of the μCT and any artefacts which may have developed during back projection). Secondly, it is dependent on the fibre tracing routine, outlined in steps 1–4 above. Both these issues must be investigated further if the 3D fibre reconstruction work is to be extended and improved.

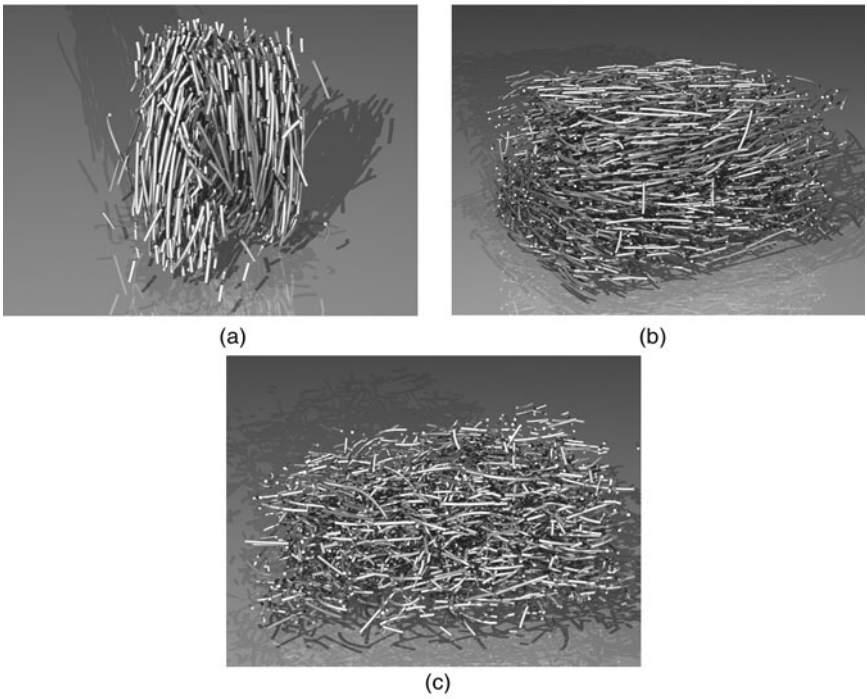


Figure 5.34 Three 3D reconstructions of different fibrous systems are shown: (a) woven yarn, (b) a fibre-reinforced composite material and (c) a non-woven textile sample.

Note that the fibre datasets, which are used to recreate the 3D views, can be used to characterise quantitatively the sample microstructure. For example, Fig. 5.35 shows part of an injection-moulded, plaque through-thickness variation in fibre orientation, which has been calculated from the 3D reconstruction data shown in Fig. 5.34(b). Just over half of the plaque thickness has been scanned. The fibre orientations are characterised by the fibre orientation tensor coefficients, a_{xx}, a_{yy} and a_{zz}, and the outer 'skin' region can be seen, clearly differentiated from the central 'core' region within the plaque.

Yet another example is the characterisation of the woven yarn sample (shown in Fig. 5.34(a)) using the fibre 'space curve' approach to determine both fibre curvature and fibre torsion. The results are shown in Figs 5.36 and 5.37. In determining the fibre curvature, only fibre lengths greater than 1 mm have contributed to the curvature determination. As far as we are aware, the frequency distribution of the fibre torsion in Fig. 5.37 is the first torsion data to have been measured directly by any experimental technique. The torsion plot was created by sampling each fibre at ten points along its length and note that there are two peaks in the distribution at ± 0.5 rad mm^{-1}. We believe that these are due to the

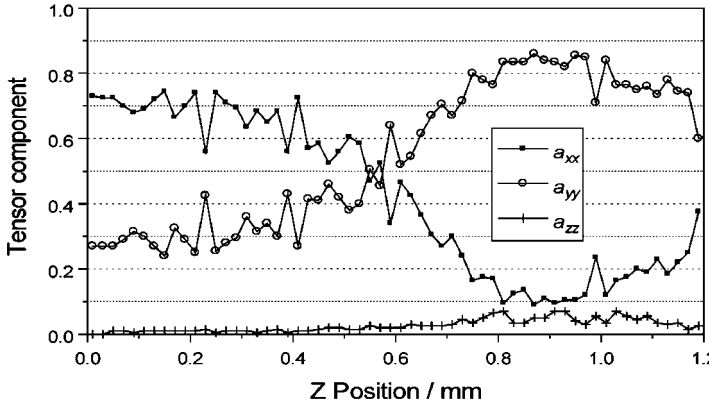


Figure 5.35 Derivation of the fibre orientation tensor coefficients a_{xx}, a_{yy} and a_{zz} from the reconstruction data presented in Fig. 5.34(b). The coefficients are plotted as a function of position through part of the thickness of a fibre-reinforced, injection-moulded sample and illustrate clearly the typical skin and core regions.

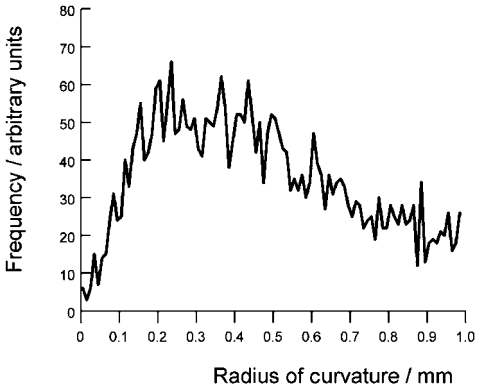


Figure 5.36 The frequency distribution of fibre curvatures for the yarn sample shown reconstructed in Fig. 5.34(a), using the space curve approach.

way that the two yarns are twisted together. The smaller, negative peak may well be due to the twist within each bundle of fibres.

5.6.8 Conclusions

The ease of sample preparation makes the X-ray technique particularly attractive to experimenters. When it does not work well with a sample, it is usually due to poor contrast between different phases within the sample (because of the phases having similar X-ray absorption spectra). In this case, it may be possible to dope interesting matrix regions during processing with, say, iodine solution at low

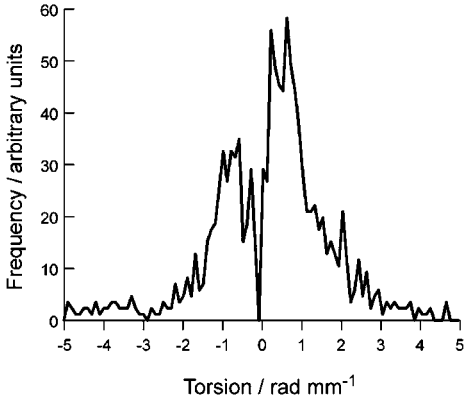


Figure 5.37 The frequency distribution of fibre torsions for the yarn sample shown reconstructed in Fig. 5.34(a), using the space curve approach.

concentrations (sufficiently low concentrations that one is confident that the dopant will not affect the flow processes).

There is a need for careful 3D comparisons to be made between CLSM and X-ray microtomography, just as we have compared CLSM to 2D image analysis in order to illustrate the errors involved with the 2D systems. By comparing confocal, sub-surface data with X-ray data, the true spatial resolution of the commercial X-ray systems could be verified and we are not aware of this cross-comparison between the two techniques having been reported by any research group.

Despite the excellent (and convincing) 3D reconstructions of complex fibrous systems, unbiased, quantitative characterisations will be dependent on the spatial resolution of the measurement system, the size of the fibres and the effects of the random noise components of the fibre centre coordinate determinations. Care must be taken when presenting these implied, 3D characterisation parameters and when critically analysing their implications.

Our latest contribution to a paper on applications of the X-ray microtomography technique may be found in Jia *et al.*[88] Finally, the X-ray microtomography technique shows great promise for materials science, especially when modifications can be made to the basic device (and which still satisfy the stringent Health and Safety requirements demanded of these machines). For example, Professor Pyrz's research team at Aalborg have built a mechanical stress stage into their Skyscan X-ray equipment to enable the real-time stressing of materials within the shielded, detection system enclosure.[89] This approach could soon give unparalleled insight into the dynamic behaviour of heterogeneous materials and yield the data necessary for improving/validating existing finite element models.

5.7 References

1 Danuser G. (2001), Super-resolution microscopy using normal flow decoding and geometric constraints, *J. Microsc-Oxford*, **204**, 136–149.
2 Gustafsson M.G.L. (2000), Surpassing the super-resolution limit by a factor of 2× using two objective lenses, *J. Microsc-Oxford*, **198**, 82–87.
3 Toraldo di Francia G. (1995), Resolving power and information, *J. Opt. Soc. Am.*, **45**, 497–501.
4 Weber R.L. (1980), *Pioneers of Science*, Institute of Physics Publishers.
5 Raman C.V. and Krishnan K.S. (1928), A new type of radiation, *Nature*, **121**, 501.
6 Smekal A. (1923), Zur Quantentheorie der Dispersion, *Naturwiss.*, **11**, 873–875.
7 Ferraro J. and Nakamoto K. (1994), *Introductory Raman Spectroscopy*, Academic Press Inc., New York.
8 Ayars E.J., Jahncke C.L., Paesler M.A. and Hallen H.D. (2000), Fundamental differences between micro- and nano-Raman spectroscopy, *J. Microsc-Oxford*, **202**, 142–147.
9 Williams K.J.P., Pitt G.D., Batchelder D.N. and Kip B.J. (1994), Confocal Raman micro-spectroscopy using a stigmatic spectrograph and CCD detector, *Appl. Spectrosc.*, **48**, 232–235.
10 Baldwin K.J., Batchelder D.N. and Webster S. (2001), Raman Microscopy: Confocal and Scanning Near-Field, Chapter 4 in *The Handbook of Raman Spectroscopy*, eds. I.R. Lewis and H.G.M. Edwards, Marcel Dekker Inc., New York.
11 Woodbury E.J. and Ng W.K. (1962), Ruby laser operation in the near IR, *Proceedings of the Institute of Radio Engineers*, **50**, 2367.
12 Turrell G. and Corset J. (eds.) (1996), *Raman Microscopy: Developments and Applications*, Academic Press, London.
13 Moskovits M. (1985), Surface-enhanced spectroscopy, *Reviews of Modern Physics*, **57**(3), Part 1, 783.
14 Betzig E., Finn P.L. and Weiner J.S. (1992), Combined shear force and nearfield optical microscopy, *Appl. Phys. Lett.*, **60**, 2484.
15 Goetz M., Drews D., Zahn D.R.T. and Wannamacher R. (1998), Near-field Raman spectroscopy of semiconductor heterostructures and CVD diamond layers, *J. Lumin.*, **76–77**, 306–309.
16 Young R.J. (1997), Analysis of composites using Raman and fluorescence microscopy – a review, *J. of Microsc-Oxford*, **185**, Part 2, 199–205.
17 Young R.J. (1996), Evaluation of composite interfaces using Raman spectroscopy, *Key Engineering Materials*, **116**, 173–192.
18 Galiotis C. (1998), Micromechanics of reinforcement using laser Raman spectroscopy, Chapter 8 in *Microstructural Characterisation of Fibre-reinforced Composites*, ed. J. Summerscales, Woodhead Publishing Ltd.
19 Binnig G., Rohrer H., Gerber C.H. and Weibel E. (1982), Surface studies by scanning tunnelling microscopy, *Phys. Rev. Lett.*, **49**, 57–61.
20 Binnig G., Quate C.F. and Gerber C.H. (1986), Atomic force microscopy, *Phys. Rev. Lett.*, **56**, 930–933.
21 Umeda N., Ishizaki S. and Uwai H. (1991), Scanning attractive force microscope using photothermal vibration, *J. Vac. Sci Tech. B*, **9**, 1318–1322.
22 Yang J., Tamm L.K., Sonlyo L.P. and Shao Z. (1993), Promises and problems of biological AFM, *J. Microsc-Oxford*, **171**, Part 2, 183–198.
23 Hu J., Xiao X.D., Ogletree D.F. and Salmeron M. (1995), Scanning polarisation

force microscopy: a technique for imaging liquids and weakly adsorbing layers, *Appl. Phys. Lett.*, **67**, 466–468.
24 Hsiao J. (1999), Correcting scan errors, *Microscopy Today*.
25 Abbe E., (1873), *Archiv f. Microskop. Anat*, **9**, 413.
26 Synge E.H. (1928), A suggested method for extending microscopic resolutions into the ultra-microscopic region, *Phil. Mag.*, **6**, 356–362.
27 Pohl D.W., Denk W. and Lanz M. (1984), Optical stethoscopy – image recording with resolution lambda/20, *Appl. Phys. Lett.*, **44**, 651–653.
28 Mariani T., Ascoli C., Baschieri P., Frediani C. and Musio A. (2001), Scanning force images through the 'Milliscope' – a probe microscope with a very wide scan range, *J. Microsc-Oxford*, **204**, Part 1, 53–60.
29 Kalkbrenner T., Ramstein M., Mlynek J. and Sandoghdar V. (2001), A single gold particle as a probe for apertureless scanning near-field optical microscopy, *J. Microsc-Oxford*, **202**, 72–76.
30 Frohn J.T., Knapp H.F. and Stemmer A. (1997), True optical resolution beyond the Rayleigh limit achieved by standing wave illumination, *PNAS*, 7232–7236.
31 Zayats A.V. and Sandoghdar V. (2001), Apertureless near-field optical microscopy via local second-harmonic generation, *J. Microsc-Oxford*, **202**, Part 1, 94–99.
32 Subramaniam V. (2001), Biological applications of multi-photon NSOM with multiple spectroscopic modes, *Microscopy & Analysis*, January, 13–14.
33 Hamann H.F., Kuno M., Gallagher A. and Nesbitt N.J. (2001), Molecular fluorescence in the vicinity of a nanoscopic probe, *J. Chem Physics*, **114**(19), 8596–8610.
34 Milner R.G. and Richards D. (2001), The role of tip plasmons in near-field Raman microscopy, *J. Microsc.*, **202**, 66–71.
35 Yamaguchi M., Sasaki Y., Sasaki H., Konada T., Horikawa Y., Ebina A., Umezawa T. and Horiguchi T. (1999), Imaging of optical disc using reflection-mode, scattering type, scanning near-field optical microscopy, *J. Microsc-Oxford*, **194**, 552–557.
36 Inouye Y. and Kawata S. (1997), Reflection-mode, near-field, optical microscope with a metallic probe tip for observing fine structures in semiconductor materials, *Opt. Commun.*, **134**, 31–35.
37 Silva T.J., Schultz S. and Weller D. (1994), Scanning near-field optical microscope for the imaging of magnetic domains in optically opaque materials, *Appl. Phys. Lett.*, **65**, 658–660.
38 Smith D.A., Webster S., Ayad M., Evans S.D., Fogherty D. and Batchelder D.N. (1995), Development of a scanning near-field optical probe for localised Raman spectroscopy, *Ultramicroscopy*, **61**, 247–252.
39 Webster S., Smith D.A. and Batchelder D.N. (1998), Raman microscopy using a scanning near-field optical probe, *Vibrat. Spectrosc.*, **18**, 51–59.
40 Webster S., Batchelder D.N. and Smith D.A. (1998), Sub-micron resolution measurement of stress in silicon by near-field Raman spectroscopy, *Appl. Phys. Lett.*, **72**, 1478–1480.
41 Bonera E., Fanciulli M. and Batchelder D.N. (2002), Raman spectroscopy for a micrometric tensorial analysis of stress in silicon, *Appl. Spec.*, **56**(5), 560–563.
42 DeWolf I., Maes H.E. and Jones S.K. (1996), Stress measurements in silicon devices through Raman spectroscopy: Bridging the gap between theory and experiment, *J. Applied Phys.*, **79**, 7148–7156.
43 Narita Y., Tadokoro T., Ikeda T., Saiki T., Mononobe S. and Ohtsu M. (1998), Near-

field Raman spectral measurement of polydiacetylene, *Appl. Spectrosc.*, **52**(9), 1141–1144.
44 Webster S., Demangeot F., Bonera E., Sands H.S., Bennett R., Hayward I.P., Marchi F., Smith D.A. and Batchelder D.N. (2002), Development of a combined confocal and scanning near-field microscope for deep UV laser excitation, in press.
45 Schultz C.P. (2001), Precision infrared spectroscopic imaging: the future of FT-IR spectroscopy, *Spectroscopy*, **16**(10).
46 Herschel W. (1800), Investigation of the powers of the prismatic colours to heat and illuminate objects, *Phil. Trans. Roy. Soc. London*, **90**, 255–283.
47 Herschel W. (1800), Experiments on the refrangibility of the invisible rays of the sun, *Phil. Trans. Roy. Soc. London*, **90**, 284–292.
48 Herschel W. (1800), Experiments on the solar, and on the terrestrial rays that occasion heat: Part I, *Phil. Trans. Roy. Soc. London*,, **90**, 293–326.
49 Herschel W. (1800), Experiments on the solar, and on the terrestrial rays that occasion heat: Part II, *Phil. Trans. Roy. Soc. London*, **90**, 437–538.
50 Birtwell N.C., Clarke A.R., Marsden P.L., Reehal J.S. and Smith C.D. (1983), Observations of large scale emission in the far infrared, *Proc. of the VIth ESA-PAC Symposium on European Rocket & Balloon Programmes & Related Research*, Interlaken, April 1983 (ESA Publication, SP-183 June 1983), pp. 435–443.
51 Fellgett P.B. (1984), 3 concepts make a million points, *Infrared Physics*, **24**(2–3), 95–98.
52 Jacquinot, P. (1984), How the search for a throughput advantage led to Fourier Transform Spectroscopy, *Infrared Physics*, **24**(2–3), 99–101.
53 Connes, P. (1986), From Newtonian fits to Wellsian heat rays – the history of multiple beam interference, *J. Opt.*, **17**(1), 5–28.
54 Tsien R.Y. and Poenie M. (1986), Fluorescence ratio imaging: a new window into intracellular ionic signalling, *Trends Biochem. Sci.*, **11**, 450–455.
55 Cole M.J., Siegel J., Webb S.E.D., Jones R., Dowling K., Dayel M.J., Parsons-Karavassilis D., French P.M.W., Lever M.J., Sucharov L.O.D., Neil M.A.A., Juskaitis R. and Wilson T. (2001), Time-domain whole-field fluorescence lifetime imaging with optical sectioning, *J. Microsc-Oxford*, **203**, Part 3, 246–257.
56 Neil M.A.A., Squire A., Juskaitis R., Bastiaens P.I.H. and Wilson T. (2000), Wide-field optically sectioning fluorescence microscopy with laser illumination, *J. Microsc-Oxford*, **197**, 1–4.
57 Carlsson K. and Liljeborg A. (1998), Simultaneous confocal lifetime imaging of multiple fluorphores using the intensity-modulated multiple-wavelength scanning (IMS) technique, *J. Microsc-Oxford*, **191**, 119–127.
58 Straub M. and Hell S.W. (1998), Fluorescence lifetime three-dimensional microscopy with picosecond precision using a multi-focal, multi-photon microscope, *Appl. Phys. Lett.*, **73**, 1769–1771.
59 Verveer P.J., Hanley Q.S., Verbeek P.W., Van Vliet L.J. and Jovin T.M. (2001), Theory of confocal fluorescence imaging in the programmable array microscope (PAM), *J. Microsc-Oxford*, **189**, Part 3, 192–198.
60 Liang M., Stehr R.L. and Krause A.W. (1997), Confocal pattern period in multiple-aperture confocal imaging systems with coherent illumination, *Opt. Lett.*, **22**, 751–753.
61 Hanley Q.S., Verveer P.J., Gemkow M.J., Arndt-Jovin D. and Jovin T.M. (1999), An optical sectioning programmable array microscope implemented with a digital

micromirror device, *J. Microsc-Oxford*, **196**, 317–331.
62 Hanley Q.S., Verveer P.J., Gemkow M.J., Arndt-Jovin D. and Jovin T.M. (2000), Three-dimensional spectral imaging by Hadamard transform spectroscopy in a programmable array microscope, *J. Microsc-Oxford*, **197**, 5–14.
63 Heintzmann R., Hanley Q.S., Arndt-Jovin D. and Jovin T.M. (2001), A dual path programmable array microscope (PAM): simultaneous acquisition of conjugate & non-conjugate images, *J. Microsc-Oxford*, **204**, Part 2, 119–137.
64 Dunkers J.P., Parnas R.S., Zimba C.G., Peterson R.C., Flynn K.M., Fujimoto J.G. and Bouma B.E. (1999), Optical coherence tomography of glass-reinforced polymer composites, *Composites Part A: Applied Science & Manufacturing*, **30**, 139–145.
65 Kenney J.M., Jacobsen C., Kirz J., Rarback H., Cinotti F., Thomlinson W., Rosser R. and Schidlovsky G. (1985), Absorption microanalysis with a scanning soft x-ray microscope: mapping and distribution of calcium in bone, *J. Microsc-Oxford*, **138**, 321–328.
66 Zhang X., Balhorn R., Mazrimas J. and Kirz J. (1996), Mapping and measuring DNA to protein ratios in mammalian sperm head by XANES imaging, *J. Struct. Biol.*, **116**, 335–344.
67 Meyer-Ilse W., Medecki H., Jochum L., Anderson E., Attwood D., Magowan C., Balhorn R., Moronne M., Rudolph D. and Schmahl G. (1995), New high resolution zone plate microscope at Beamline 6.1 of the ALS, *Synchrotron Radiation News*, **8**, 29–33.
68 Meyer-Ilse W., Denbeaux G., Johnson L.E., Bates W., Lucero A. and Anderson E. (1999), The high resolution X-ray microscope, XM1, *X-ray Microscopy: Proc. 5th Int. Conf. on X-ray Microscopy*, Berkeley August 2–6 (ed. Meyer-Ilse, Warwick & Attwood) pp 129–134. AIP Conf Proc 507, American Institute of Physics, College Park.
69 Niemann B., Rudolph D. and Schmahl G. (1983), The Gottingen X-ray microscopes, *Nucl. Inst. Methods*, **208**, 367–371.
70 Ice G. and Bai J.M. (2000), *www.ornl.gov/news*, as reported in 'Letting the grains show their strains', *Materials World*, May, 20–21.
71 Berglund M., Rymell L., Peuker M., Wilhein T. and Hertz H.M. (2000), Compact water-window transmission X-ray microscopy, *J. Microsc-Oxford*, **197**, Part 3, 268–273.
72 Darlington M.W., McGinley P.L. and Smith G.R. (1976), Structure and anisotropy of stiffness in glass fibre-reinforced thermoplastics, *J. Mat. Sci.*, **11**, 877–886.
73 Geier M.H. (1994), *Quality Handbook for Composite Materials*, Chapman & Hall Publishers, pp. 225–232.
74 Maisl M., Scherer T., Reiter H. and Hirsekorn S. (1988), Nondestructive investigation of fibre reinforced composites by x-ray computed tomography, in *Nondestructive Characterisation of Materials*, eds P. Holler, V. Hauk, G. Dobmann, C. Ruud and R. Green, Springer-Verlag, pp 147–154.
75 Awaji M., Suzuki Y., Takeuchi A., Takano H., Kamijo N., Tamura S. and Yasumoto M. (2001), X-ray imaging microscopy at 25 keV with Fresnel zone plate optics, *Nucl. Instrum. Meth. A.*, **467**(2), 845–848.
76 Herman G.T. (1980), *Image Reconstruction from Projections: the Fundamentals of Computed Tomography*, Academic Press, New York.
77 Shepp L.A. and Vardi Y. (1982), Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging*, **1**, 113–122.

78 Feldkamp L.A., Davic L.C. and Kress J.W. (1984), Practical cone-beam algorithm, *J. Opt. Soc. Am.*, **1**(6), 612–619.
79 Hsieh J. (1995), Image artifacts, causes, and correction, in *Medical CT and Ultrasound: current technology and applications*, eds. L.W. Goldman and J.B. Fowlkes. Advanced Medical Publishing, Madison.
80 Gong R.H. and Newton A. (1996), Image-analysis techniques Part II: The measurement of fibre orientation in nonwoven fabrics, *J. Text. I.*, **87**(2), 371–388.
81 Pourdehyhimi B., Dent R., Jerbi A., Tanaka S. and Deshpande A. (1999), Measuring fiber orientation in nonwovens, Part V: Real webs, *Text. Res J.*, **69**(3), 185–192.
82 Xu B.G. and Yu L. (1997), Determining fiber orientation distribution in nonwovens with Hough transform techniques, *Text. Res. J.*, **67**(8), 563–571.
83 Tao X.M. (1996), Mechanical properties of a migrating fibre, *Text. Res. J.*, **66**(12), 754–762.
84 Huh Y., Kim Y.R. and Ryu W.Y. (2001), Three-dimensional analysis of migration and staple yarn structure, *Text Res J.*, **71**(1), 81–90.
85 Papathanasiou T.D. and Guell D.C. (eds.) (1997), *Flow-induced Alignment in Composite Materials*, Woodhead Publishing Ltd., Cambridge.
86 Kang T.J. and Lee S.H. (1999), Characterization of reinforcing web structures in needle punched nonwoven composites, *J. Comp. Mater.*, **33**(22), 2116–2132.
87 Eberhardt C.N. and Clarke A.R. (2002), Automated reconstruction of curvilinear fibres from 3D datasets acquired by X-ray microtomography, *J. Microsc-Oxford*, **206**, Part 1, 41–53.
88 Jia X., Gopinathan N., Williams R.A., Eberhardt C.N. and Clarke A.R. (2001), X-ray Microtomography Facilitated Modelling of Microstructures, *Proc. 2nd World Congress on Industrial Tomography*, Hannover, Germany, August, pp. 451–460.
89 Pyrz R. (1999), Application of X-ray microtomography to the study of polymer composites, *Proceedings of the ICCM12*, Paris, July, p. 479.

6

Other microscopy techniques

> To probe the structure of atoms, Rutherford suggested to his graduate students, Geiger and Marsden, an investigation on the scattering of energetic alpha particles by thin metal foils. When they reported that one particle in 8000 was deflected by more than a right angle, Rutherford is reported to have remarked later 'it was as if you had fired a 15-inch naval shell at a piece of tissue paper and the shell came right back and hit you!'
>
> ... from *Pioneers of Science*, R.L.Weber

There are basically three reasons why the microscopist may be considering non-optical techniques in research:

(1) to achieve higher spatial resolutions than optical systems
(2) to probe the internal structure of materials, which are opaque at optical wavelengths
(3) to determine physical parameters, which cannot be measured by electromagnetic radiation.

The spatial resolution achievable by different experimental techniques has been illustrated in Chapter 1. Rayleigh's qualitative criterion for the ultimate limit to the spatial resolution, discussed in Chapter 1, showed that better resolutions could be achieved only by using smaller wavelength probes (due to the diffraction limit imposed by microscope lenses and other apertures). Hence, spatial resolutions may be improved by using radiation in the ultraviolet or X-ray region of the EM spectrum.

However, in the early days of the development of Quantum Theory, when de Broglie suggested that all elementary particles with momentum, p, had an associated wave-like property of wavelength, λ, given by

$$\lambda = \frac{h}{p} \qquad [6.1]$$

where h is Planck's constant, it was soon realised that a new type of particle microscope could be developed. If the trajectories of high energy particles could

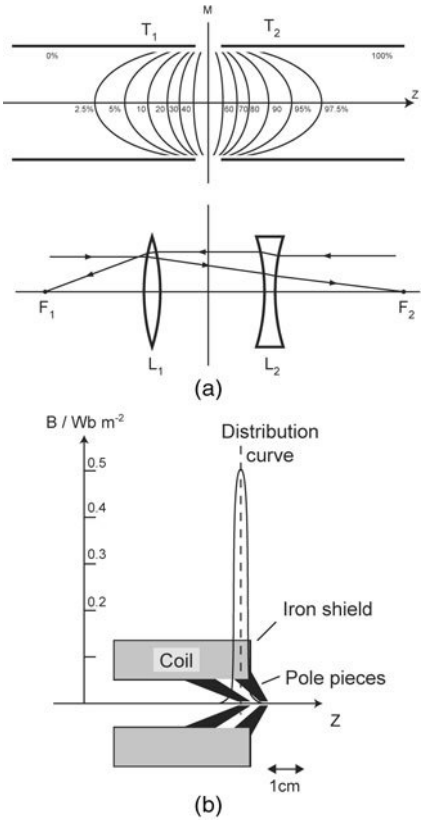


Figure 6.1 Basis of the use of magnetic and electrostatic lenses for electron beam refraction and focussing. (a) Different electric potentials on two cylindrical tubes create the equi-potential curves shown. Electrons moving through the tubes, close to the z-axis, will be deflected in the same way that light rays would be refracted through (and focussed by) the combination of the convex lens, L_1 and the concave lens, L_2. (b) The alternative method of focussing electrons is by winding a coil, surrounding it with an iron shield and arranging for specially shaped pole pieces to concentrate the magnetic field, when a current flows through the coil, as shown. This configuration produces, in effect, a short focal length lens.

be focussed like light rays, then the spatial resolutions achievable should be given by Rayleigh's criterion using de Broglie's wavelengths. In electron microscopes, electron trajectories are manipulated by electrostatic and magnetic fields and are focussed onto specimens, just as glass lenses refract the light waves in optical microscopes (see Figs 6.1 and 6.2). The interaction of the electrons with the sample (which is usually placed in an evacuated chamber to prevent attenuation of the electron beams by scattering off air molecules) gives nanometre scale size, spatial resolutions.[1,2]

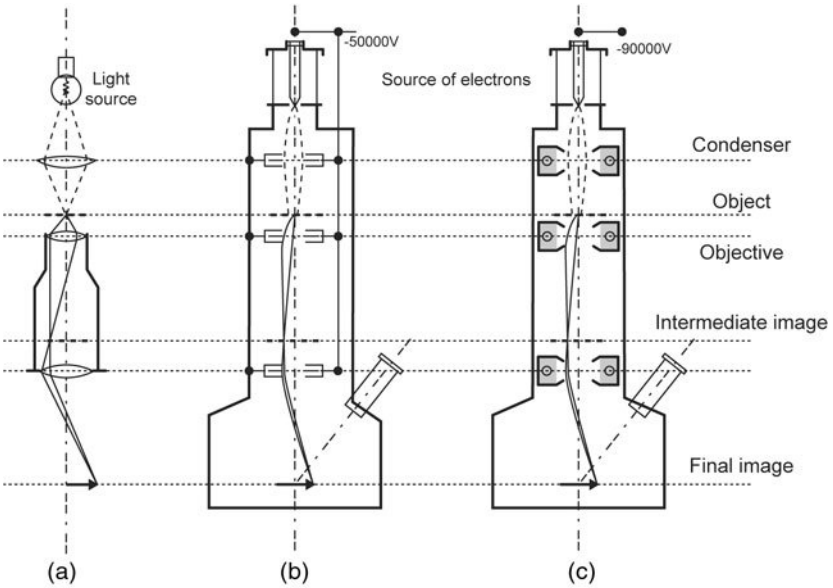


Figure 6.2 Comparison of the basic operation of (a) a standard optical microscope, (b) an electrostatic version of the electron microscope and (c) a magnetic coil version of the electron microscope.

Nowadays, the highest spatial resolutions (at atomic scale sizes) are being achieved by the new scanning probe techniques; see, for example, the review by Grim and Hadziioannou.[3] In 1986, Ernst Ruska received the Nobel Prize in Physics for his work on the development of electron microscopy together with the German scientists, Gert Binnig and Heinrich Rohrer, who in 1981 developed the scanning tunnelling microscope (STM) technique.[4,5] These microscopic techniques, discussed in Chapter 5, will have an increasing role to play in cutting-edge research in the near future with the latest engineering developments and novel applications in nanotechnology. However, the latest electron microscopes have a greater depth of field (around 10 μm) at these high resolutions than the STMs and hence have considerable use in microscale and mesoscale research within materials science.

Whereas the scanning probe microscopy techniques essentially explore the surface of a specimen, other techniques allow the user to look for different inhomogeneities in the bulk of the specimen. For example, neutrons, X-rays and electrons are complementary probes employed to study the structure of materials because neutrons interact with nuclei, X-rays with electrons in the materials, and incident electrons with electrostatic potentials. As the cross-section for absorption of X-rays and neutrons is lower than for electrons, far greater penetration depths may be achieved in materials science with X-ray and neutron techniques.

6.1 Electron microscopy (SEM/TEM)

There are two types of electron microscope – the scanning electron microscope (SEM) and the scanning transmission electron microscope (STEM). The first commercial electron microscope was manufactured by Cambridge Instruments in 1965 and is shown in Fig. 6.3. A modern commercial electron microscope, manufactured by Philips is shown in Fig. 6.4. These are high cost devices and all require measurements to be undertaken in a vacuum, except for the environmental SEM discussed later. Also, the samples must be conducting (in order to accelerate the electrons onto the sample) and, hence, a biological sample must have a gold layer deposited on its surface if it is to be investigated by SEM or STEM. In the STEM, the sample is a very thin specimen and contrast within the image is due to the spatial variations in intensity of the transmitted electron beam through the specimen, as the beam is raster scanned over the specimen. In the SEM, the image may be produced in a number of ways – from variations in the intensity of secondary electrons back-scattered from the specimen through to X-ray emission produced by inelastic collisions of the primary beam with bound electrons in the specimen.

The idea that gave rise to the electron microscope is that, just as light is refracted and focussed by an optical lens, the electron, due to its charge, will

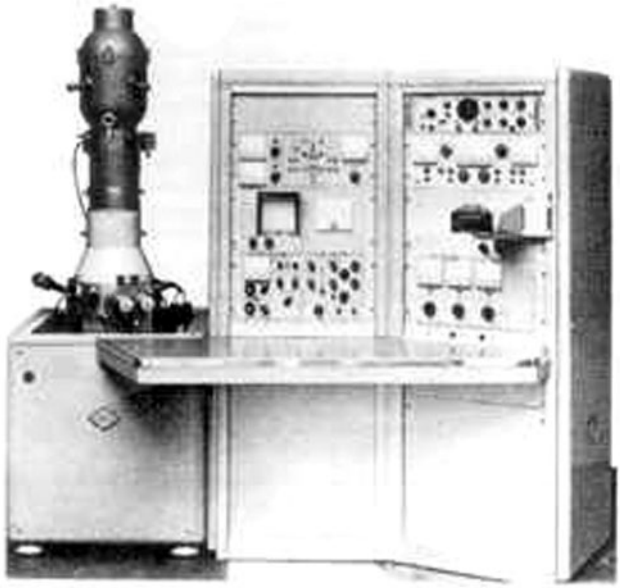


Figure 6.3 Photograph of the very first commercial electron microscope produced by Cambridge Instruments Ltd (courtesy of Leica).



Figure 6.4 Photograph of one the most modern, futuristic looking, scanning electron microscopes, which is manufactured by Philips.

have its path deviated by either a magnetic field or an electric field. Therefore, with careful design of appropriate electric and magnetic fields within the instrument, an electron beam may be focussed like an optical beam. A schematic view of the STEM configuration is shown in Fig. 6.2(b) and (c). The electrons are emitted by an incandescent cathode source, accelerated towards more positive grids through either electrostatic or magnetic field lens onto an object. The specimen is supported on a very thin film to minimise the scattering of the electrons as they pass through the sample. Depending on the thickness and composition of the object, the electron beam experiences different attenuation as a function of position. The beam travels through two more lenses before being imaged onto a fluorescent screen (in original models) or photographic plate or directly onto a scintillator placed on the face of a photomultiplier tube or a CCD device. (A scintillator is a semi-transparent material, which emits a flash of light when a charged particle traverses it.) The spatial resolution of this type of microscope is determined by the wavelength associated with the electrons and this wavelength may be 100,000 times smaller than optical wavelengths at the typical accelerating voltages used in electron microscopy.

TEM in a high-resolution (nanometre resolution) mode can be used with thin ultra-microtomed specimens 40–60 nm thick. Fine structural details smaller than 1 nanometre will be detectable. In Fig. 6.2 (b) and (c), the schematic diagram of a transmission electron microscope system was shown. The parallel illuminating electron beam is scattered by the object. The scattered beam passes through the convergent lens, L, and is focussed in the intermediate image focal plane as an electron diffraction pattern. The final optical image is formed in the Gauss plane. For the magnified image to be a reliable Fourier synthesis, all of the beams must interfere whilst keeping their initial relative phases. The objective lens will introduce additional phase shifts. These phase shifts vary as a function of the spherical aberration C_s of the lens, the diffraction angle, 2θ, of the beams relative to the optic axis and the defocus distance, Δf. The transfer function of the electron beam lens, K, has been shown to be[6]

$$K(u) = D(\lambda f u)e^{-i\chi(u)} \qquad [6.2]$$

where

$$\chi(u) = \pi(\tfrac{1}{2}C_s\lambda^3u^4 - \lambda\Delta f u^2) \qquad [6.3]$$

and D is the characteristic function of the objective aperture, $\chi(u)$ is the aberration function, u is the spatial frequency $(u = 2\theta/\lambda)$, f is the focal image distance and λ is the wavelength. The objective acts like a filter for spatial frequencies, selecting and modifying the phase of planar waves that can interfere in the image plane.

Typical acceleration voltages are $V = 120\,\text{kV}$, which gives the electron an energy, $E = 120\,\text{keV}$ (where 1 electronVolt, eV is equivalent to an energy of 1.6×10^{-19} Joules) and the representative wavelength, λ, for these electrons is 3.35×10^{-3}nm, computed from the non-relativistic De Broglie expression

$$\lambda = \frac{h}{\sqrt{2Em_e}} \qquad [6.4]$$

where m_e is the electronic mass. Hence the electron microscope has potentially, one of the highest spatial resolutions of all of the microscope techniques discussed in this book. For example, a recent article by Den Dekker *et al.* discusses how to optimise the design of a quantitative, high resolution, electron microscope (HREM) to attain the highest spatial resolution.[7] Den Dekker's argument is that the traditional qualitative Rayleigh criterion for point source resolution may have been a convenient rule of thumb for qualitative microscopy but is not appropriate for the latest computer-aided, quantitative microscopies. The electron microscope (like any other microscope) can be considered as an information channel that carries information from object to observer.[8] In order to quantitatively extract as much information as possible, one should use a model for the overall imaging process, including

(i) the electron-object interaction
(ii) the microscope transfer function and
(iii) the image detection (by CCD or whatever sensing device).

The overall model for the electron microscope contains parameters, e.g. atomic coordinates which are unknown and which must be determined by experiment. Structure determination can be achieved by fitting the model to the experimental data in an iterative procedure in which the model parameters are systematically refined. By implication, detail beyond the Rayleigh point resolution (and even beyond the information limit of the microscope) may be extracted because, in the absence of noise (and assuming that the model is correct), the fitting procedure would be perfect and exact values of the model parameters would be obtained! In practice, however, results will always be limited by some noise component, e.g. the Poisson noise due to the counting statistics (quantum or shot noise) which is determined by the number of electrons in the illuminating beam. Den Dekker's paper discusses the precision with which a single atom's position could be estimated from noisy HREM images.

In scanning electron microscopes, low energy secondary electrons ($E_e < 50\,\text{eV}$) are detected using a scintillator/pm tube detector. At these low energies, the secondary electrons can only escape, and be detected, from within a few nanometres of the surface of the specimen. Note that the SEM has a similar spatial resolution to the atomic force microscope (AFM), which was discussed in detail in Chapter 5. So one might be prompted to ask 'What are the relative advantages/disadvantages of these two types of microscopes for nanometre scale size materials science research?'.

Consider investigating the surface structure of thin films. SEM has the capability of imaging a large area (several mms by several mms) whereas the AFM typically images $100\,\mu$m by $100\,\mu$m areas. However, changes in the surface slope can cause higher SEM intensities with an ambiguity in the sense of the surface gradient whereas AFM detects surface gradients unambiguously. In the semiconductor industry, these microscopies are used to investigate high aspect ratio structures (i.e. surfaces with well-defined trenches and holes). If the SEM is to be used, the samples are usually cleaved and their cross-section is imaged from end-on. The AFM, on the other hand, can be used without destroying the specimen, but care has to be taken with the probe tip shape and, in general, if the vertical structures show an 'over-hang' effect, the AFM probe will not be able to detect it. Finally, if the specimen has a rough surface (showing hundreds of microns through to millimetres of vertical information), it can be imaged with the SEM (because of its large depth of field) whereas the AFM can only cope with vertical features around 5–10 μm.

Traditionally, the electron microscopes have been operating with the specimen in a high vacuum (whereas the AFM is usually used in standard laboratory conditions or in a liquid). The need for a high vacuum is a serious

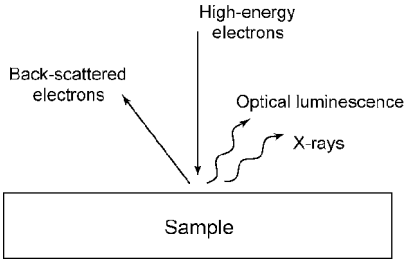


Figure 6.5 Schematic view of the different products created after an electron beam hits a material, e.g. back-scattered electrons, X-rays, cathodo-luminescence.

handicap for many investigations and hence the more recent development of the environmental scanning electron microscope (ESEM) (see for example Danilatos,[9] Howie,[10] Thiel *et al.*[11] and Stelmashenko *et al.*[12]), has been a significant one. Although it achieves a lower spatial resolution than the original vacuum SEMs, the ESEM has extended electron microscopy into novel application areas.

Figure 6.5 shows the various types of electrons and EM radiation produced when a high energy electron beam strikes the specimen. Most electron microscopes have extra attachments to allow them to generate and analyse X-rays or generate and analyse the energy spectra of the secondary emission electrons. These attachments create a more powerful tool for the study of chemical constituents of surfaces. There is a plethora of acronyms for the different measurement techniques associated with electron microscopy. For example, researchers use energy dispersive X-ray emissions (EDX), wavelength dispersive X-ray emissions (WDX), electron back-scattered diffraction (EBSD), electron beam induced currents (EBIC), X-ray photoelectron spectroscopy (XPS) and Auger electron spectroscopy (AES). Only a few of these techniques will be mentioned in the next few subsections and interested readers are advised to visit the electron microscope manufacturers websites, for example Oxford Instruments and Leica,[13] to find out more specific information on these techniques.

6.1.1 XPS/ESCA/AES techniques

There is one technique which has made a big impact on surface science and especially the acquiring of analytical chemical information about different elements at sample surfaces. The technique is sometimes called X-ray photoelectron spectroscopy (XPS) and, at other times, electron spectroscopy for chemical analysis (ESCA).[14] It depends on the photo-ionisation produced when a high energy photon (in the X-ray region of the EM spectrum) strikes a

surface. Typically, the sample is placed in a high vacuum chamber and a monochromatic beam of soft X-rays or electrons irradiates the sample. (Usually the magnesium K_α or the aluminium K_α X-ray lines are selected.) The beam of X-rays (or electrons) produce photo-ionisation and the kinetic energy of the ejected photoelectrons is measured by passing the electrons through some kind of energy analyser (either a time-of-flight or magnetic analyser or electrostatic energy analyser). The equation which describes the energetics of the photo-ionisation process is

$$h\nu = E_B + E_K \qquad [6.5]$$

where $h\nu$ is the incident X-ray photon (or electron) energy, E_B is the binding (ionisation energy) of the electron from a particular energy level, and E_K is the kinetic energy of the ejected photoelectron. Hence, a measurement of the kinetic energy of the photoelectron allows for the determination of the binding energy and hence elemental identification and chemical information on the surface materials. A typical output consists of a plot of electron counts (proportional to the intensity) as a function of electron binding energy or of electron kinetic energy, as shown in Fig. 6.6(a). The interpretation of this plot, which was generated by a magnesium K_α source onto a glass fibre-reinforced composite, is given by the energy level diagrams in Fig. 6.6(b) and (c).

The peaks in the spectrum are indicative of oxygen and carbon, and Auger peaks arising from these elements are noted also. The in-coming X-ray photon energy (or electron beam energy) is more than sufficient to release electrons from the 1s, 2s and 2p levels of both carbon and oxygen. The Auger process is a secondary ionisation event where the 'hole' (created by the loss of an electron through primary ionisation) captures an electron from an upper energy level. The electron ionised from the upper level is the Auger electron and another acronym often seen is AES, which stands for Auger electron spectroscopy. Also, one might see X-AES for X-ray induced AES or the alternative, electron induced AES. For the applications of scanning Auger microscopy to materials science, see Paparazzo who points out that all elements except for hydrogen and helium can be identified with the AES technique because of their characteristic bound electron energy levels.[15] Wider applications of scanning Auger microscopy are also discussed in Ferguson.[16]

An attractive feature of XPS/ESCA is its extreme surface sensitivity. Only electrons that are not inelastically scattered between the point of formation and the electron detector yield these peaks in the energy spectra. If they are inelastically scattered, they contribute to the background counts. Typical electron escape depths (high probability of contributing to the spectral peaks) are in the range 0.5–3 nanometres and hence this represents the effective depth of penetration of this technique. For more details of the application of this technique to composite materials, see Dillard.[14]

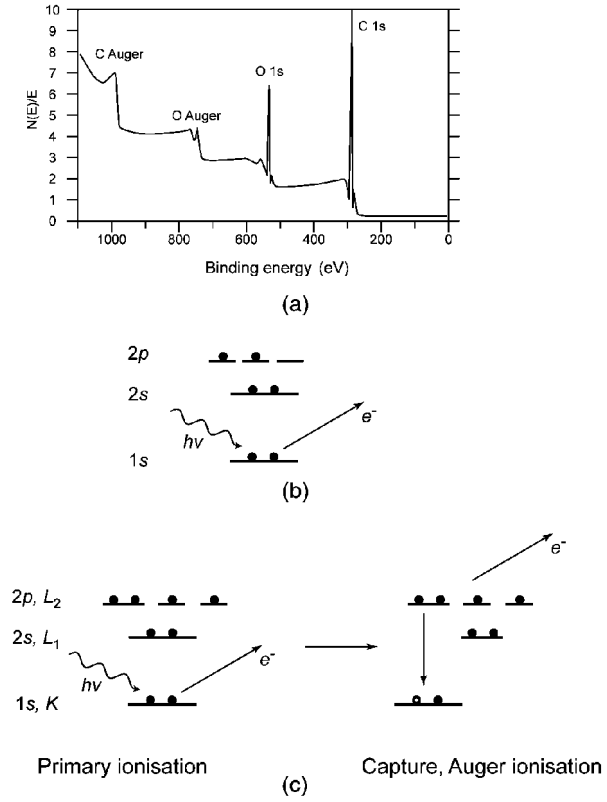


Figure 6.6 (a) Typical XPS/ESCA spectra produced by a polyester-based glass fibre-reinforced composite. (b) Schematic of the photo-ionisation process giving ionisation of the C 1s electron and (c) the schematic of the oxygen KLL Auger process.

6.1.2 Back-scattered electrons and the EELS technique

Back-scattered electrons are those electrons in the incident electron beam, which experience a series of elastic scattering events in the specimen that completely reverses their original direction of travel and causes them to be emitted from the top surface of the specimen. Experimenters can explore the energy and angular distributions of these back-scattered electrons, their depth of generation and lateral spread, their incident beam energy dependency and their specimen tilt dependency. Figure 6.7 shows the energy distribution relative to the incident beam energy of electrons emitted from a typical solid specimen. Note that higher atomic number target materials yield higher numbers of energetic back-scattered electrons. Hence, this effect forms the basis of compositional imaging (characterising regions of specific elements at the specimen surface) using back-scattered electrons. One application area has been the analysis of the

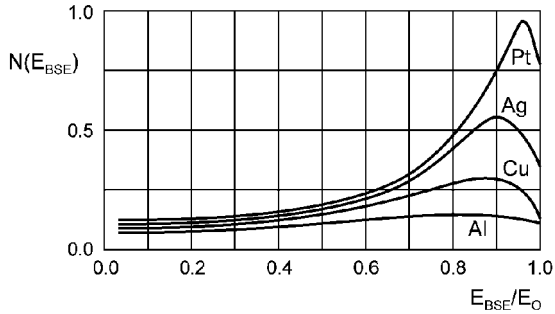


Figure 6.7 The energy spectra of back-scattered electrons from different high Z target materials. It is clear that a significant fraction of the back-scattered electrons have energies close to the initiating electrons.

mineral distribution in bone samples,[17] where the calcium content can be distinguished successfully from the lighter elements carbon, oxygen and nitrogen, which compose the organic fraction of bone.

Research continues on simulating the back-scattered process by Monte Carlo methods in order to gain a better insight into this measurement technique.[18, 19] Also, an Internet-accessible database has recently been established, containing thousands of entries on electron-matter interactions and electron scattering cross-sections, as described in Napchan,[20] which aims to help researchers in this area.

A similar technique is electron energy loss spectroscopy (EELS).[21] Using the classical picture of the scattering of fast electrons by an atom, the incident electron sees a nucleus that is screened by an electron cloud. Hence, the scattering amplitude is determined mainly by the net charge of the ion at small scattering angles (i.e. the spatial distribution of the electrostatic potential is strongly influenced by ionicity and charge transfer). Therefore, with an advanced electron microscope, it should be possible to reveal the spatial distribution of charge within nanometre-sized volumes, in combination with studies of the symmetry and density of electronic states above the Fermi level through the fine structure exhibited near to the absorption edge in electron energy-loss spectroscopy.

The full schematic diagram of the STEM, used by Terauchi *et al.* for EELS, is shown in Fig. 6.8.[21] Note the monochromator action of the Wien filter (in crossed electric, *E*, and magnetic fields, *B*, an electron's path will not be deflected if the electron velocity is equal to the ratio, *E/B*). Hence, an energy selector can be made which allows electrons of a particular velocity (and hence energy) to pass through. Terauchi's group describe the complex design of a successful Wien filter, which enables off-axis electrons (with velocity *E/B*) to be selected too. The improved energy sensitivity achieved with their design (see Fig. 6.9), has been used for investigating optoelectronic materials like GaN

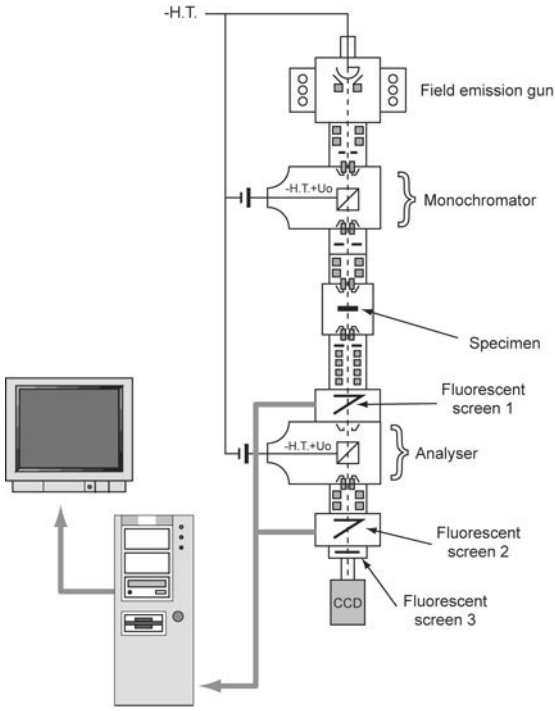


Figure 6.8 Schematic diagram of the high energy resolution EELS microscope, adapted from Terauchi *et al.*[21] Both monochromator and analyser contain Wien filters, which use specially designed arrangements of orthogonal electric and magnetic fields to give this microscope its unique properties.

(gallium nitride). Other EELS research descriptions can be found in Leapman[22] and Lakner.[23]

Space does not permit describing other variants on the electron microscope measurement technique in any detail. Some of the techniques seen in the recent literature are:

- *High angular annular darkfield (HAADF) imaging* – this technique has been used to explore the compositional profiles of monolayers of dopants in semiconductors, which exhibit special optical and electrical properties.[24]
- *Lorentz microscopy and electron holography* – this technique has been used to investigate the magnetic microstructure of a variety of nano-crystalline materials. De Graef *et al.* used a Philips CM200-FEG transmission electron microscope.[25] An electrostatic bi-prism enabled them to perform off-axis electron holography. Images and holograms were recorded digitally using a 1024×1024 pixel multi-scan CCD camera. A bi-prism voltage of 90 volts was used with a magnification of $70{,}000\times$ to give a 343 nm field of view and a pixel size of 2.7 nm.

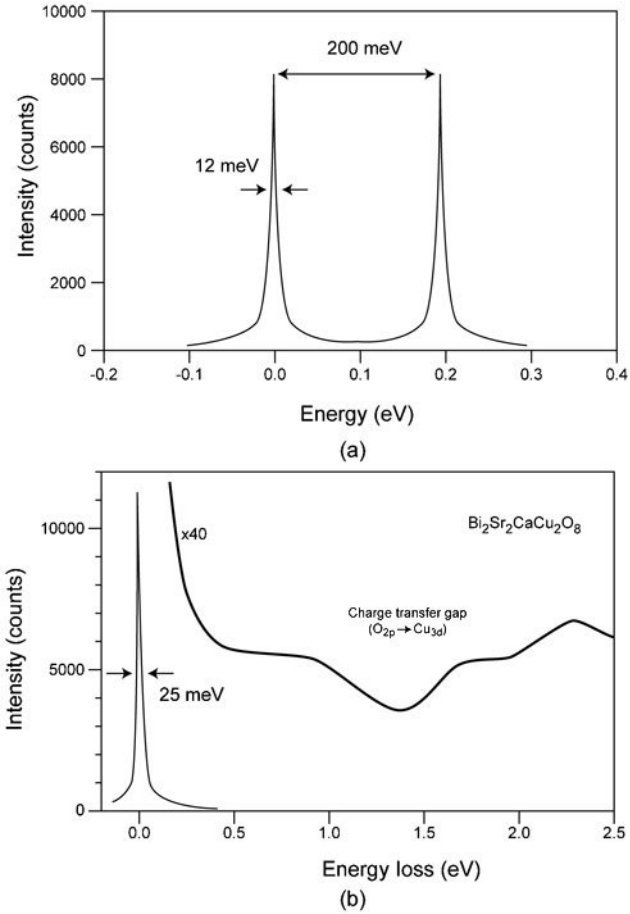


Figure 6.9 (a) The ultimate energy sensitivity of the EELS system is illustrated by this plot showing the energy spread of the zero-loss beam without a specimen in place. The FWHM of the peak was calculated to be 12 meV by referring it to the peak whose energy difference was known to be 200 meV. (b) The electron energy-loss spectrum of an oxide superconductor of $Bi_2Sr_2CaCu_2O_8$ obtained with an energy resolution of 25 meV. Both plots have been adapted from Terauchi *et al.*[20]

- *Energy filtering transmission electron microscopy (EFTEM)* – this technique has been used for the creation of elemental maps by selecting electrons with a specific energy.[26]
- *Convergent beam electron diffraction (CBED)* – this technique has been investigated for deducing strain in novel semiconductors by the shift of Bragg lines. A new 'rocking curve' analysis technique is described in Zipprich *et al.*[27]

which leads to local lattice strains being measured to high accuracy and high spatial resolution.

- *Electron beam induced current (EBIC)* – this technique has been used to determine the minority carrier diffusion lengths in semiconductors. This measurement tool enables the researcher to identify lateral inhomogeneities in diffusion lengths and also the diffusion length depth distribution may be deduced using a depletion region modulation technique.[28]

6.2 Nuclear magnetic resonance (NMR)

There are a number of excellent books on all aspects of nuclear magnetic resonance (NMR), including Callaghan[29] and Koenig.[30] What follows is merely an indication of the process behind the technique, spatial resolutions achievable and typical material studies that have been undertaken by NMR.

6.2.1 Basic principles

It is a well-known fact that the energy levels of an atom are split when in the presence of an external magnetic field (the Zeeman effect) because of the interaction of the atomic magnetic moment and the field. Since nuclei have magnetic moments too, the energy of a nucleus is also split in the presence of a magnetic field. Consider the simplest case of the hydrogen atom in an external field. The potential energy, U, of a magnetic moment, μ, in an external field, $\boldsymbol{B}$, is given by

$$U = -\mu \cdot \boldsymbol{B} \qquad [6.6]$$

where μ and $\boldsymbol{B}$ are vector quantities and the symbol '$\cdot$' denotes the scalar product of these two vectors. (The magnitude of the scalar product of the two vectors is $\mu B\cos\theta$, where θ is the angle between the two vectors, as shown in Fig. 6.10(a).) Therefore, the potential energy is lowest when the magnetic moment is aligned with the field and highest when it is in the opposite direction to the field. Since the spin quantum number of the proton is $\frac{1}{2}$, the proton's magnetic moment has two possible orientations in an external magnetic field: parallel to the field (spin up) or anti-parallel to the field (spin down). The difference in energy of these two possible orientations, shown in Fig. 6.10(b), is therefore given by

$$\Delta E = 2(\mu_z)_p B \qquad [6.7]$$

where $(\mu_z)_p$ is the proton moment (approximately $2.8eh/4\pi m_p = 8.79 \times 10^{-8}$ eV/T). When hydrogen atoms are irradiated with photons of energy ΔE, some of the nuclei are induced to make transitions from the lower state to the higher state by resonance absorption. These nuclei then decay back to the lower energy state, emitting photons of energy ΔE. The frequency of the photons absorbed and emitted is calculated from $\Delta E = h\nu$. A simple calculation

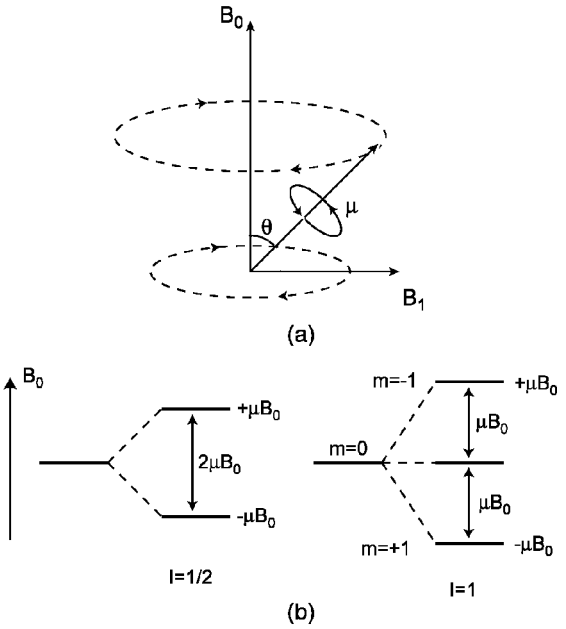


Figure 6.10 (a) When placed in a magnetic field, B_0, nuclei having spin will precess about the magnetic field direction at the Larmor precession frequency. If a second magnetic field, B_1, is introduced, which rotates at the Larmor frequency, it will cause large oscillations in the angle between μ and B_0, causing a resonance effect. (b) From a quantum mechanics viewpoint, there is a separation between magnetic energy levels (Zeeman levels) and certain transitions between levels are allowed (depending on the nuclei spin value, I), which result in the emission of a radio frequency photon.

shows that the frequency of these photons is $\nu = 42.5$ MHz (in other words, it is radio frequency EM radiation) in a 1 Tesla magnetic field. By measuring the exact radio frequency emitted, the total internal magnetic field in the molecule may be estimated (the sum of the external magnetic field and the local field due to the atoms and nuclei of the surrounding material). This is the principle behind the nuclear magnetic resonance (NMR) technique. The sample is irradiated by a broadband radio frequency source and the emitted radio frequency photons are detected. Figure 6.11 shows a photograph of a typical commercial NMR system.

6.2.2 Practical considerations

NMR is based on the fact that many atomic nuclei oscillate like tiny gyroscopes when in a magnetic field. In NMR, a sample is placed in a magnetic field, which forces the nuclei into alignment. The sample is bombarded with radio waves and, as the nuclei absorb the radio waves, they topple out of alignment with the

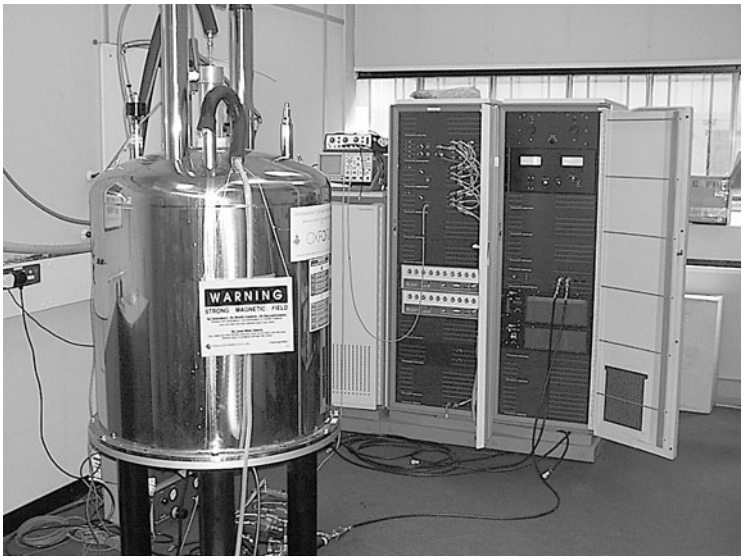


Figure 6.11 A typical nuclear magnetic resonance research system used in the Department of Physics and Astronomy at the University of Leeds. The cryostat ensures that the sample for analysis is kept at sub-zero temperatures.

magnetic field. Hence, the nuclei lose the energy absorbed from the radio waves and re-align themselves again to the magnetic field. By measuring the specific radio frequencies that are emitted by the nuclei and the rate at which the re-alignment occurs, the spectroscopists can obtain detailed information about the molecular structure and the molecular motion within the sample being studied. NMR imaging is a technique to create maps, which relies on placing the sample in a spatially inhomogeneous magnetic field whose nuclear resonance frequency is matched to the radio frequency signal in a small region.

The NMR signals depend on the nuclear relaxation time constants T_1 and T_2, which reflect the structural environment of the emitting nucleus. There are several ways to spatially encode the NMR signal. One way is to apply a linear magnetic field to the original static field, creating a magnetic field strength gradient across the sample. In 1D, the position of the sample, z, is related to a frequency, ω_z, by the relationship

$$\Delta\omega_z = \omega_z - \omega_o = \gamma \frac{\partial B_z}{\partial z} z \qquad [6.8]$$

where γ is a constant called the gyromagnetic ratio for hydrogen. A radio frequency pulse (with a narrow frequency range) is used to excite only those nuclei at corresponding positions in the z-direction. The amplitude of the NMR signal received from the z-axis line is a measure of the number of resonant nuclei along that line.

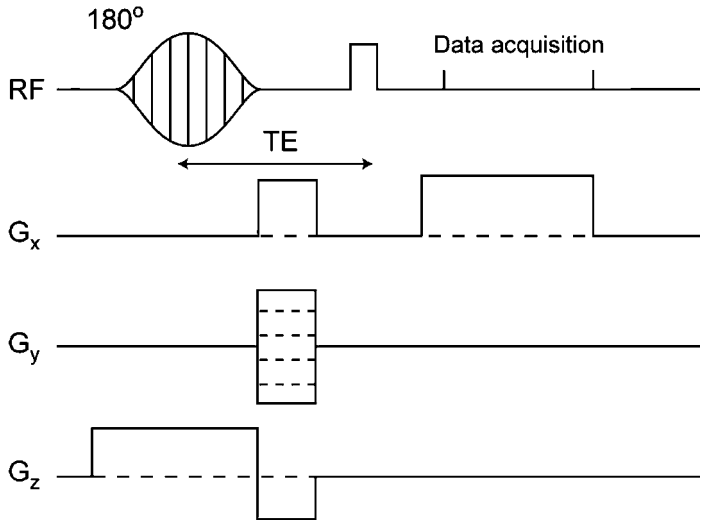


Figure 6.12 A timing diagram for the signals which implement the NMR spin-echo scheme – see the description in the main text.

The spin densities and the molecular environments of the nuclei are reflected in the variation in the amplitude of the measured radio frequency (rf) signal and hence in the voxel intensities. When the time constants T_1 and T_2 are different in heterogeneous samples, these differences may be exploited to give contrast within the image. The pulse sequence, which achieves this image contrast enhancement, is called 'multiple spin echo' and the process is illustrated in Fig. 6.12. In the spin echo method, a series of rf pulses are emitted (to improve the overall signal-to-noise ratio by averaging over the series). Each projection is produced by a 90° pulse, followed by a 180° pulse for induction of the spin echo. The 90° pulse tips the magnetisation into the XY plane, where it begins dephasing. The 180° rf pulse is applied after a time, t, and forces the magnetisation to refocus at time $2t = \mathrm{TE}$, the echo time, after the 90° pulse. Data are then collected. The frequency-encoding gradient G_x causes the spins to precess at different frequencies depending on their position relative to the static magnetic field. The phase encoding gradient, G_y, is orthogonal to G_x. Varying the phase of G_y causes the spins to dephase at different rates, providing the second dimension of a 2D image. The slice-selection gradient, G_z, and the Gaussian-shaped 90° rf pulse determine the position and thickness of the region of interest. The data are Fourier transformed in 2D to produce the image of a selected slice. The time between two pulse sequences is usually around 250 milliseconds through to 2500 milliseconds and hence quite long data collection times are necessary. Also note that the spin-lattice relaxation time, T_1, will limit the minimum data collection time.

Spatially resolving a given volume in an NMR image is equivalent to performing NMR spectroscopy on that volume. Two spatially distinct volume elements will be resolved if the magnetic field gradient is of sufficient strength that the elements are shifted in resonance frequency from each other by an amount greater than the natural line width of a spectral line. The spatial resolution, Δx, may therefore be written

$$\Delta x = \frac{\delta\omega}{\gamma\dfrac{\partial B_x}{\partial x}} \qquad [6.9]$$

where $\delta\omega$ is the line width. For mobile liquids, the line widths are very narrow (giving high spatial resolution) and volumes of the order of $10\,\mu\text{m} \times 10\,\mu\text{m} \times 100\,\mu\text{m}$ may be distinguished. Routine measurements of liquids in voids or cracks within solids or liquid-solid mixtures have typically $40\,\mu\text{m} \times 40\,\mu\text{m} \times 100\,\mu\text{m}$ volume resolutions.

The difficulties of solid state imaging are due to the solid state line width being 1000 times broader than the solution counterpart! Therefore, either the magnetic field gradient has to be increased by a factor of 1000, or some other cunning means has to be found to reduce the overall data acquisition times. These techniques are beyond the scope of this book and readers should consult texts such as Koenig.[30]

6.2.3 NMR applications to materials science

Two recent reviews of the applications of NMR to composites and polymer research have been compiled by Matzkanin[31] and Blumich and Blumler.[32] Nuclear magnetic resonance (NMR) has been used in these research areas to locate localised defects, to image local stress distributions associated with filler agglomerates and to image diffusion and non-Newtonian flow in polymer melts.

NMR microscopy achieves NMR imaging at spatial resolutions δx, δy and $\delta z \geq 10\,\mu\text{m}$, which is a much poorer spatial resolution than for optical microscopy. However, the use of larger field gradients and multi-pulse sequences should improve this NMR spatial resolution. NMR imaging has been applied to composites with some success but it has one major drawback. The imaging of glassy polymers such as epoxy is difficult because of the long spin-lattice relaxation time (T_1) that leads to line broadening and very short spin-spin relaxation time (T_2) that cannot be detected with current electronics. Hence, samples are usually infused with a liquid and it is the relaxations of the liquid that are monitored. For example, voidage within pultruded rods has been investigated by soaking the rods in warm water for months and then using the spin-echo technique to pick out water-filled voids, as described in Hoh *et al.*[33]

Carbon fibre composites can be imaged (except where the plane of the laminate sheets is perpendicular to the radio frequency field, because the

conductive sheets screen the radio frequency field within the coil), see Jezzard *et al.*[34] For example, 1 mm is the greatest practical depth for imaging a carbon fibre cross-ply composite when the NMR fields are oriented perpendicular to the fibres, whereas the practical limit for imaging is 3 cm when the NMR field orientation is parallel to the fibre direction. Matzkanin[31] presents 55 references to the applications of NMR for composites research and shows how NMR can be used to look at product degradation, internal strains, environmental stress cracking and blending of polymers through to studying polymer chain dynamics (using carbon-13 and deuterium as probes).

6.3 Ultrasonics and scanning acoustic microscopy

What follows is a very short account of acoustic microscopy, but so broad is this area that interested readers should consult one of the many excellent books on this subject, e.g. Briggs.[35]

6.3.1 Introduction

A totally different type of wave to electromagnetic waves is the acoustic or pressure wave, which was touched on in Chapter 1. In homogeneous materials, the velocity of different pressure waves, c_a, depends on the appropriate elastic constant, K, and the density of the material, ρ.

$$c_a = \sqrt{\frac{K}{\rho}} \qquad [6.10]$$

For instance, the Young's modulus, Y, or the shear modulus, S, are appropriate elastic constants for longitudinal and transverse waves respectively in solids, and the bulk modulus, B, is appropriate in the case of longitudinal sound waves through gases. Hence the velocities of these pressure waves in solids is in the range 1600–2500 m/sec, in liquids it is in the range 900–1600 m/sec and in gases like air the velocity is around 340 m/sec. (All pressure wave velocities are dependent on the ambient temperature.)

Spatial resolution in a conventional microscope depends on the wavelength of the waves illuminating the specimen. It is interesting to note that, as these pressure wave velocities are much lower than the velocity of light, the frequencies needed to give wavelengths in the microns range are also much lower than for light. Hence, scanning acoustic microscopes (SAM) operate at frequencies around 1–2 GHz to produce images with conventional optical resolutions, whereas low frequency ultrasound (in the 2–10 MHz range) gives spatial resolutions in the millimetre range. Low frequency ultrasound can be used to probe the spatial variations in elastic constants throughout a specimen, as shown in the Case Study in Section 6.4. Unfortunately, as the frequency increases, the depth of penetration decreases dramatically because the

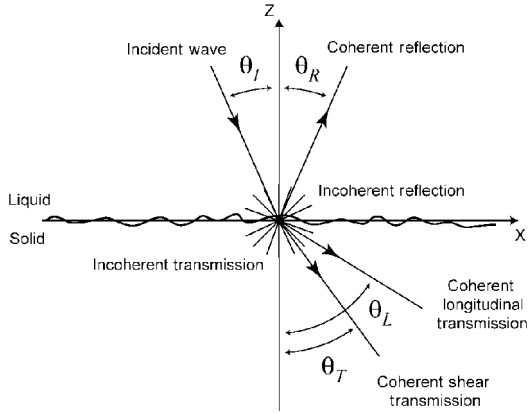


Figure 6.13 A schematic diagram of the various acoustic waves which could exist at a liquid/solid interface, when the solid exhibits surface roughness.

attenuation of the pressure waves increases as the square of the frequency. Hence there is a trade-off between resolution and specimen penetration. These pressure waves, regardless of frequency, are more heavily attenuated in air than in liquids and so water is usually used as a convenient couplant between the transmitter of the pressure waves and the specimen and also the specimen and the receiver.

Gases and liquids only support longitudinal wave modes, but solids support the propagation of transverse (shear) waves too. When an acoustic wave strikes an interface between say liquid and solid, there are a number of effects. Some of the energy is reflected back into the liquid, some of the energy goes into the transmission of quasi-transverse and quasi-longitudinal waves (see Fig. 6.13), and also a very small fraction (10^{-5}) of the incident energy may create surface acoustic waves (SAW). There are two possible types of shear wave that appear in the solution to the wave equation. However, the longitudinal waves in water can only couple into vertically-polarised shear waves in the solid (not horizontally polarised shear waves, which would be polarised in the plane of the interface).

Low frequency continuous pressure waves can be produced by small-scale loudspeakers (after all, these are the sources of music in the range 100–20,000 Hz). The higher frequency waves are usually excited by pressure transducers, which are often made of piezoelectric materials and which have low mass and hence high resonance frequencies. Most types of ultrasonic systems use pulsed ultrasound – a high voltage, short duration pulse excites the transducer and the transducer produces a few waves at its resonant frequency.

Over the past few decades, the explosion of interest in these pressure wave techniques has spawned almost as wide a range of variations as for optical microscopy – all the way from tomographic techniques,[36] through to combining

ultrasound with atomic force microscopes, see Rabe *et al.*[37] Rabe *et al.* have applied an alternating electric field between a conducting cantilever and a piezoelectric sample and, through the inverse-piezoelectric effect, have set the sample surface into vibration. This vibration was localised around the contact area formed by the sensor tip and the sample surface.

6.3.2 Scanning acoustic microscopy

In 1936, S.Y. Sokolov (in USSR Patent No. 49) first described the possibility of utilising acoustic waves for microscopy and his ideas were extended by Dunn and Fry in 1959 at the University of Illinois.[38] Eventually, it was the development of the pulsed acoustic systems (which enabled the reflections of the acoustic beam from the specimen to be separated from spurious reflections) and the creation of images by raster scanning the specimen under a focussed beam that have led to the designs available today.

A typical tilted acoustic wave source and associated 'objective lens' for acoustic microscopy is described in Briggs[35] and is shown in Fig. 6.14. Because of the problem of severe attenuation of high frequency acoustic waves in air, the ZnO (zinc oxide) piezoelectric transducer is bonded to the top surface of a fused quartz (or sapphire) block. When this transducer is excited, plane waves are generated, which propagate down towards the concave-shaped bottom surface of

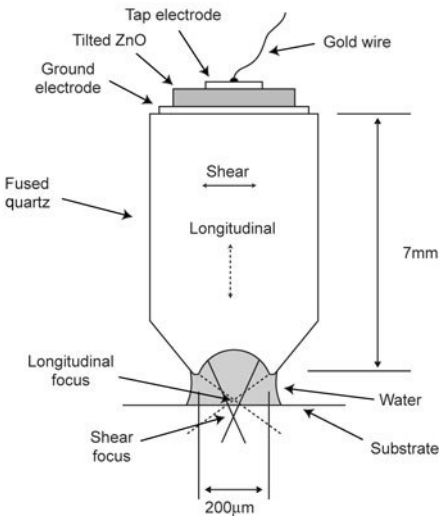


Figure 6.14 The ZnO transducer and associated fused quartz (or sapphire) lens of a scanning acoustic microscopy (SAM) system are shown. The tilted transducer allows for both tensile and shear waves to be launched into the lens. These systems operate at very high frequencies (hundreds of MHz or a few GHz).

the sapphire block. The concave surface is coated with a quarter wave thick matching film (to minimise any wave reflection at this surface – see Chapter 1) and the surface focusses the acoustic waves onto the specimen via a coupling fluid (usually water).

Transmission acoustic microscopes have been built but there are difficulties in aligning both transmitting transducer plus objective and the receiving transducer plus lens. It is far more common to arrange for one transducer to act as both transmitter and receiver of the acoustic waves. This is achieved by pulsing the transducer and then switching it to 'receive mode' ready for the echo from the specimen.

In an acoustic microscope, 2D images are built up by scanning a point transducer in X and Y. The same aspects of the material structure which influence the propagation of these mechanical (pressure) waves determine the stiffness and mechanical performance of the specimen. Either one must probe the surface structure with focussed high frequency (i.e. low wavelength) pressure waves as in scanning acoustic microscopy (SAM) to achieve the highest spatial resolution or, at lower frequencies, obtain greater depths of penetration into the material. Note that 1 GHz frequencies have spatial resolutions of a few microns, but low penetration depth, whereas frequencies around a few MHz have millimetre wavelengths, but significant depths of penetration (many cms). The problem with high frequency operation is that the bandwidth of the electronic system must be correspondingly wider and hence extra noise will be introduced – which limits the resolution. Scanning acoustic microscopy (SAM) has made significant strides over the past few years. Transducers operating at frequencies, $\nu \leq 200$ MHz and capable of focussing the ultrasonic waves may be used to scan samples in 3D with δx, δy and $\delta z \geq 20\,\mu$m spatial resolutions; see, for example, Lisy *et al.*[38] Good contrast can be achieved but other image artefacts seem to appear which would make fully-automated, 3D operation extremely difficult. Such artefacts include the presence of interference fringes around fibre images below the surface, due to surface acoustic wave reflections from the matrix-fibre interface or micro-cracks in the matrix. Acoustic microscopy can image different phases of a composite matrix (see, for example, Fatkin *et al.*[40]).

6.3.3 Ultrasonic 'time-of-flight' scanning systems

Whereas SAMs have to operate at very high frequencies (to achieve the similar spatial resolution as optical microscopes), another type of design involves time-of-flight systems. In these systems, spatial resolution is compromised in order to build up a picture of elastic constant variations in processed parts, for example the Leeds TOF system, which will be described in detail in Section 6.4.

Traditionally, there are a number of ways of achieving pulsed, time-of-flight, ultrasonic operation, as shown in Fig. 6.15. In some systems, the transducer used for the generation of ultrasound is also used as the receiver of the ultrasound

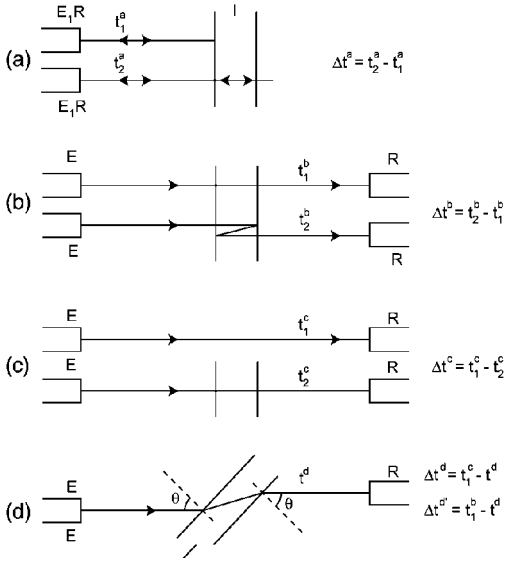


Figure 6.15 Different schemes are shown which utilise reflections in pulsed ultrasonic measurement systems. (a) A single transducer is used to transmit and receive, and in (b), (c) and (d) two transducers are employed.

echo. As the echo delay time is dependent on the distance of the reflecting object into the sample, time-gated methods may be employed to select the effective 'focal plane' of these systems within the sample. These techniques have been used extensively for the non-destructive testing of very large composite parts for evidence of voidage, and spatial resolutions, Δz, at the 20–50 μm level are common with these systems. Also, by Fourier transforming the time profiles of both transmitted and received pulses, any scattering of the ultrasonic waves by small voids within the sample may be detected through wavelength dependent attenuation, which varies as λ^{-4}, because of the Rayleigh scattering process; see, for example, de Belleval *et al.*[40]

6.4 Case study: ultrasonic mapping of 3D stiffness constants of composite materials

6.4.1 Introduction

In previous case studies (see Chapters 3, 4 and 5), the background to composites research has been given and the need for high quality microstructural data has been presented. The performance of most fibre-reinforced composite components in real applications depends on a knowledge of their elastic constants (or engineering moduli like Young's modulus, shear modulus and Poisson's ratio). Even if the modelling of the links between processing and

microstructure have not been determined with precision, a knowledge of the elastic constants (or the equivalent stiffness constants) of the component part is most valuable. At Leeds, it was decided to investigate the links between fibre spatial distributions and these stiffness constants by utilising low frequency ultrasonic waves and various optical techniques for the determination of fibre orientations. Even the latest theoretical treatises on acoustic wave propagation through materials have to make assumptions when predicting wave motion through heterogeneous materials like composites; see, for example, Nayfeh.[42]

One can describe the deformation of a solid under an applied extensional force by Hooke's Law where the stress, σ (or extensional force per unit area), the resulting strain, ε (or change in length per unit length), and Young's modulus, Y, are related by

$$Y = \frac{\sigma}{\varepsilon} \qquad [6.11]$$

In general, a complex solid in equilibrium, which is subject to a set of applied forces (P_1, P_2 and P_3), will experience many components of stress, as shown in Fig. 6.16. If the body is in equilibrium, then $\sigma_{ij} = \sigma_{ji}$ and so all similar sub-volumes of a real part will have the stress within that sub-volume defined by a nine-component matrix, with six independent components

$$\sigma_{ij} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \qquad [6.12]$$

These forces in equilibrium will cause deformations within the sub-volume and the strain components, ε_{ij}, when defined by

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \qquad [6.13]$$

give a similar strain matrix,

$$\varepsilon_{ij} = \begin{pmatrix} e_{xx} & \frac{1}{2}e_{xy} & \frac{1}{2}e_{xz} \\ \frac{1}{2}e_{xy} & e_{yy} & \frac{1}{2}e_{yz} \\ \frac{1}{2}e_{xz} & \frac{1}{2}e_{yz} & e_{zz} \end{pmatrix} \qquad [6.14]$$

where

$$e_{xx} = \frac{\partial u}{\partial x}, e_{yy} = \frac{\partial \nu}{\partial y}, e_{zz} = \frac{\partial w}{\partial z}, e_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}, e_{xz} = \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z},$$
$$e_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \qquad [6.15]$$

Hence the generalised form of Hooke's Law relating stress to strain may be written in the short-hand notation with the coefficients, c_{ijkl} called the stiffness constants of the material

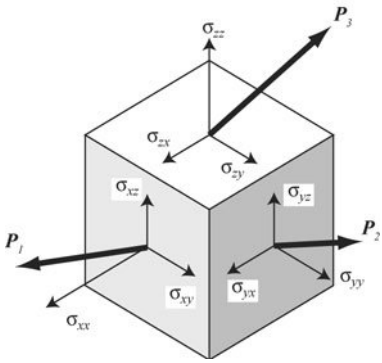


Figure 6.16 Resolved 3D stresses, σ_{ij}, on a small volume element within the sample, due to the applied forces, P_1, P_2 and P_3 on the three faces of the volume element.

$$\sigma_{ij} = c_{ijkl}\varepsilon_{kl} = \sum_{kl} c_{ijkl}\varepsilon_{kl} \qquad [6.16]$$

This defines a stiffness matrix with 81 components but, because there are only six independent components of both stress and strain, a simplified, 36 component stiffness matrix can be defined with stiffness constants, C_{pq}, as follows. Note that there are only 21 different stiffness constants due to symmetry

$$C_{pq} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{pmatrix} \qquad [6.17]$$

For example, a typical relationship between stress and strain can now be written

$$\sigma_{xx} = C_{11}e_{xx} + C_{12}e_{yy} + C_{13}e_{zz} + C_{14}e_{xz} + C_{15}e_{yz} + C_{16}e_{xy} \qquad [6.18]$$

and similarly for the other σ_{ij} coefficients.

If there is no interaction between the extensional force and the shear force, a simplification can be made by assuming that the stresses and strains are defined parallel to the coordinate axes, which are the symmetry axes of the material. It can be shown that further simplifications can be made for specific types of material,[43] e.g. a polymer which is a uniaxially oriented sheet is 'transversely isotropic' and, in this case, the number of independent stiffness constants reduces still further to six components.

$$C_{pq} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix} \quad [6.19]$$

where

$$C_{55} = C_{44} \text{ and } C_{66} = \frac{C_{11} - C_{12}}{2} \quad [6.20]$$

By considering the continuity equation throughout a small sub-volume, the link between the stiffness constants and ultrasonic velocities is established via the equation

$$\frac{\partial \sigma_{ij}}{\partial x} = \rho \frac{\partial^2 u_i}{\partial t^2} \quad [6.21]$$

where ρ is the density of the material. Substituting for σ_{ij} gives

$$c_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l} = \rho \frac{\partial^2 u_i}{\partial t^2} \quad [6.22]$$

This equation is reminiscent of the standard form of the wave equation shown in Chapter 1. There are three possible wave solutions for these equations: one wave has its displacement vector lying parallel to the direction of motion (a longitudinal wave solution) and the other two waves have orthogonal transverse displacement vectors (two polarised shear wave solutions). The relationship between the longitudinal wave velocity, V_L, and the stiffness constants and the transverse wave velocities, V_S, and the stiffness constants is given on page 390. Surface acoustic waves (SAW) may also be excited at the interface, but they are evanescent waves, which exist in the vicinity of the interface and they do not propagate energy through the material; see for example Briggs.[35]

6.4.2 Methodology

Another complementary measurement technique, which characterises the effect of 3D microstructure on global mechanical properties, is the ultrasonic, time-of-flight, immersion technique, as described by Markham[44] and Read and Dean.[45] The ultrasonic measurements may be correlated to 3D confocal (and 2D) fibre orientation state measurements in order to test models which relate microstructure to material properties. The first manual instrument was made in Leeds over 15 years ago,[46] and has been in constant use in the Polymer Physics Group; see Hine *et al.*[47] Although this instrument has been very successful, its operation was tedious. For the past six years, a fully automated version of this device has

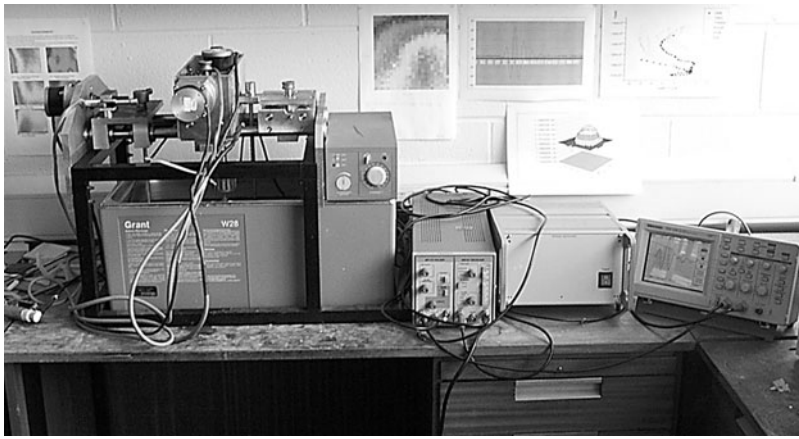


Figure 6.17 A photograph is shown of the Leeds ultrasonic, immersion testrig, which consists of a temperature-regulated water bath, stepper motor controlled sample holder, signal conditioning and computer interfacing electronics and Tektronix digital oscilloscope. The controlling host computer is off picture.

been built by members of the Molecular Physics and Instrumentation Group at Leeds. This time-of-flight, 'water immersion' testrig has been described in Enderby *et al.*[48] and, more recently, in Harper and Clarke[49] (see Figs 6.17 and 6.18).

The automated ultrasonic immersion testrig is a combination of off-the-shelf commercial subsystems and in-house electronic interfaces and mechanical construction. Initially a sub-nanosecond timer circuit was designed and constructed at Leeds. However, this has now been replaced by the timing circuitry inside the Tektronix TDS220 digital storage oscilloscope. This special, economical oscilloscope performs the timing, pulse data capture and formats the digitised dataset for transmission to the host computer through a high speed GPIB serial link. The transmitter and receiver are I3-0208-R immersion transducers from Diagnostic Sonar Ltd. (One pair of transducers was selected for an operating frequency of 2.25 MHz and another pair for 5 MHz.) Metrotek modular ultrasonic instrumentation units generate the initiating pulse (an MP215 spike pulser) and condition the received pulse group signals (an MR101B receiver/amplifier).

The thermostatically controlled water bath is a Grant W28 unit and a total of four stepper motors are under the control of the host computer via a standard RS432 serial link and a STAMP microprocessor controller. One stepper motor moves the transmitter transducer towards or away from the receiver transducer. (This function is needed to determine the velocity of the ultrasound in water *before* the sample is placed in the tank.) A second stepper motor moves the sample horizontally along a path perpendicular to the ultrasonic beam axis between the transmitter and receiver. (This function enables us to create a

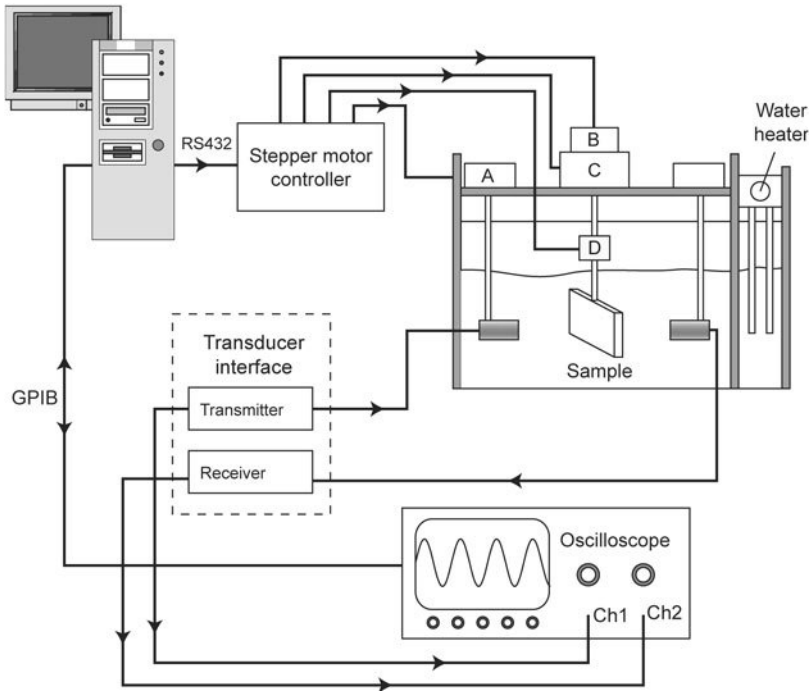


Figure 6.18 Overall schematic diagram of the Leeds ultrasonic TOF testrig, which shows the inter-connections between the different hardware components and the flow of control signals and data acquired.

horizontal line scan by moving across the beam.) A third stepper motor is capable of rotating the sample through 360º in 0.9º steps, in order to measure the variation in received signal as a function of the angle of incidence of the beam to the sample surface. The fourth stepper motor is able to raise or lower the sample, enabling the automatic raster scanning of the sample within the fixed ultrasonic beam and hence the determination of 2D maps of time-of-flight and amplitude data.

The Tektronix storage oscilloscope measures the time delay between the initiating pulse and the start of the 1024 sample digital recording of the received pulse profile (see Fig. 6.19). By measuring this time delay between the transmitter and receiver to sub-nanosecond accuracy, one can determine the velocity of a 2.25 MHz ultrasonic pulse through the water and then through the water and sample.

When the ultrasonic beam strikes the specimen at normal incidence, the longitudinal (or tensile) wave velocity, V_L, is given by

$$\frac{1}{V_L} = \frac{1}{V_W} - \frac{\Delta t}{d} \qquad [6.23]$$

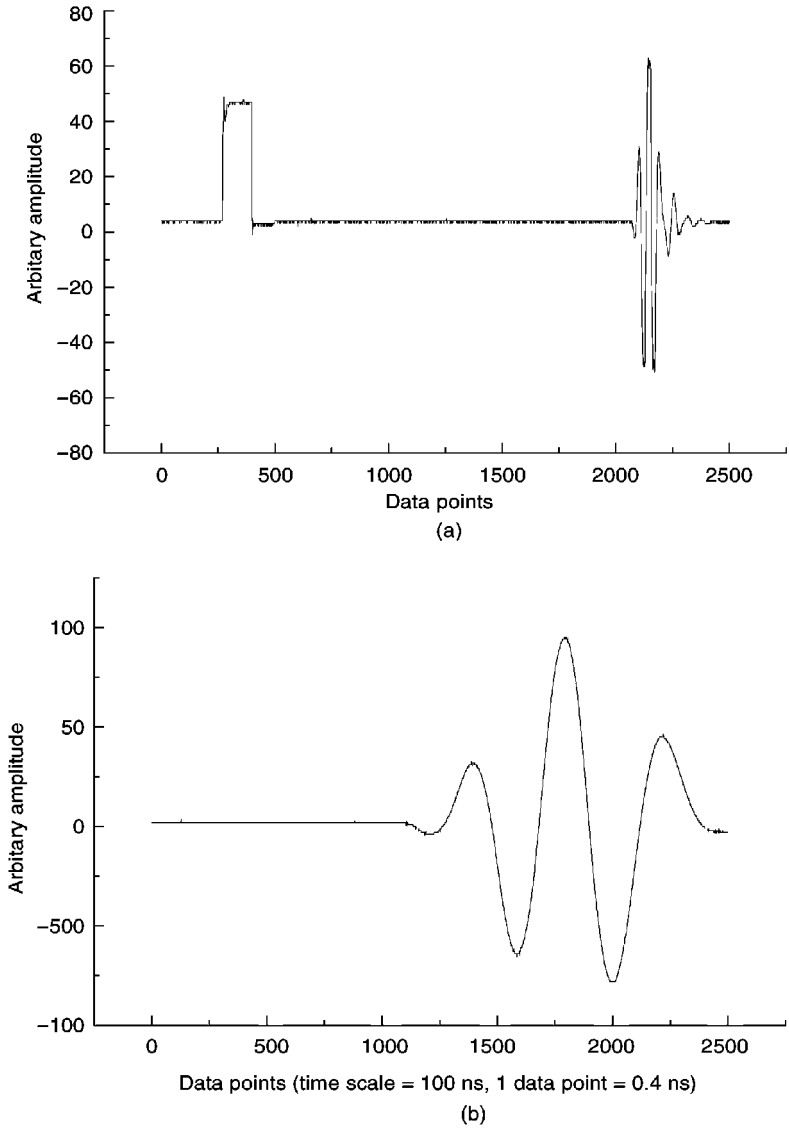


Figure 6.19 (a) Transmitted and received pulse trains are shown. (b) Magnified view of the signal received with offset time, where the time resolution (determined by the storage oscilloscope circuitry) is 0.4 nsec.

where V_W is the velocity of the longitudinal wave in water, Δt is the difference between TOF recorded with and without the sample and d is the thickness of the sample.

For non-normal incidence, at an angle i, as shown in Fig. 6.20, the TOF, Δt is dependent on the angle of refraction r, of the wave through the specimen

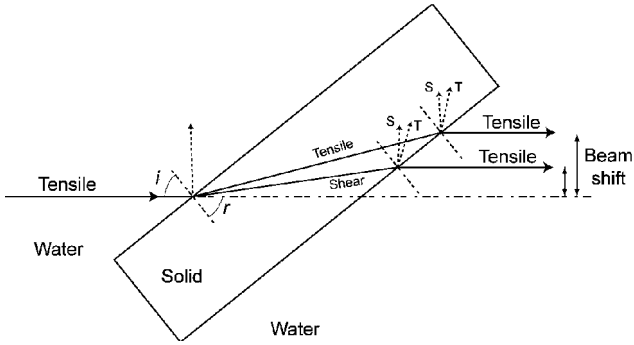


Figure 6.20 A schematic diagram is shown of the angles of incidence, reflection and refraction for ultrasonic waves propagating through a parallel-sided plaque in water. Note that the emerging longitudinal waves in the water (due to the quasi-longitudinal and quasi-transverse waves in the sample) are displaced from the original path of the incident wave.

$$\tan r = \frac{\sin i}{\cos i + q} \qquad [6.24]$$

where

$$q = \frac{\Delta t}{d} V_W \qquad [6.25]$$

Hence the shear wave velocity, V_S, for a wave travelling at an angle of refraction, r, may be found from Snell's Law

$$V_S = V_W \frac{\sin r}{\sin i} \qquad [6.26]$$

Note that there will be a similar equation relating the tensile wave velocity, V_L, for this angle of incidence.

The procedure therefore is, firstly, to measure the velocity in water and maintain the temperature of the water bath constant. As illustrated in Fig. 6.21, there is a significant change in the variation in the tensile (and also the shear) wave velocity with water temperature, which could give a serious systematic error in the interpretation of stiffness constants.

If the sample surface is normal to the ultrasonic beam path, one stepper motor can be used to scan the sample through the beam. In this way, linear scans may be constructed of the straight-through, tensile velocity component or linear scans at any fixed angle of incidence. A typical stretch of data is shown in Fig. 6.22 for a glass fibre-reinforced, polymer composite. Note the clear spatial structure in this 'time-of-flight' plot. Because the computer records a pulse group time delay and a maximum received voltage signal (and hence the ultrasonic pulse amplitude) once every millisecond (i.e. the ultrasonic pulse repetition rate is 1 millisecond), there are two important consequences:

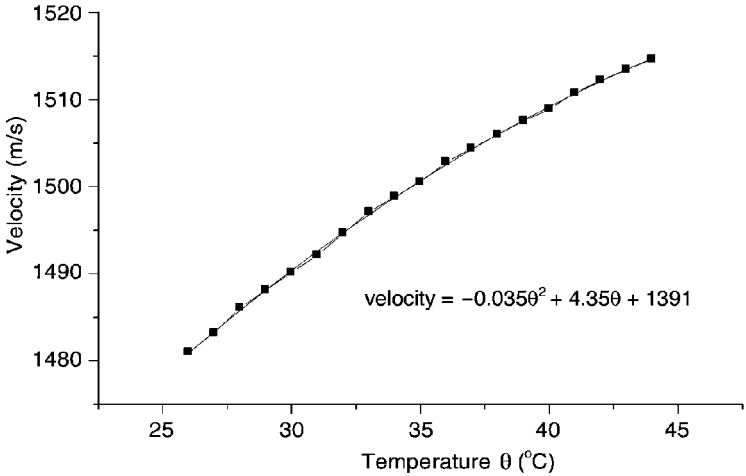


Figure 6.21 The variation of the 2.25 MHz, ultrasonic wave velocity as a function of the temperature in the waterbath, illustrating the need for temperature stabilisation when TOF measurements are acquired over a long period.

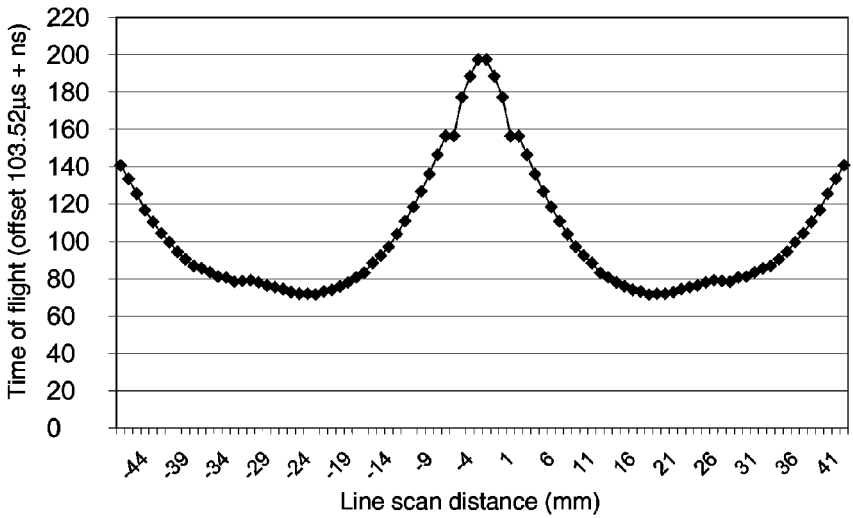


Figure 6.22 A typical linescan of the time-of-flight (TOF) of a normal incidence, ultrasonic wave as a fibre-reinforced composite sample is scanned forwards and then backwards across the fixed beam path. Note the symmetry of the curve about the zero position, which is taken as the end of the forward scan. The reproduceability of the spatially-varying ultrasonic TOF data is clear.

(a) The system integrates over many measurements at the same settings and, in effect, achieves sub-nanosecond accuracy in estimating the ultrasonic pulse time delays, and improves the pulse profile in preparation for Fourier transforming of the received pulse, if necessary.
(b) The system can scan very quickly and, after scanning forwards and backwards, remove any systematic temperature-dependent effects or minimise noise (which may be due to external mechanical vibrations).

By incorporating another stepper motor, the linear (X) motion of the sample through the beam is combined with a perpendicular motion (Y) to create 2D maps of time-of-flight and amplitude. These maps give the operator a quick overview of interesting regions within the sample. (Note, however, that a set of similar 2D maps, at different angles of incidence, must be generated in order to deduce the variation of stiffness constants in this sample region, as described below.)

6.4.3 Perspex investigation/calibration

Like any other wave propagation, the amplitude of the received ultrasonic pulse will be determined by (i) the transmission coefficient through the first interface, (ii) the attenuation of the wave as it propagates through the sample and (iii) the transmission coefficient at the second interface. The attenuation of the 2.25 MHz ultrasound pulse in Perspex, as shown in Fig. 6.23, was obtained by transmitting the ultrasound through different thickness Perspex samples at normal incidence. This water immersion system cannot measure the transmission coefficients at single interfaces, but the received signal is, in effect, giving information on the overall transmission through the liquid/solid and solid/liquid interfaces.

A typical time-of-flight 2D map for Perspex (at normal incidence) is shown in Fig. 6.24. In order to search for trends, the time-of-flight and the amplitude values for each corresponding ‘pixel’ in the 2D maps are plotted in the cross-correlation graph of Fig. 6.25. Clearly, in this typical, parallel-sided, isotropic sample, there is only a small residual correlation, which is most likely due to systematic variations in the sample thickness across the area scanned.

However, the system has another powerful mode of operation. The sample may be rotated by a third stepper motor with respect to the ultrasonic beam direction. In this way, the angle of incidence is varied, and, if one plots the variation in amplitude of the transmitted beam, through the Perspex block, the graph shown in Fig. 6.26(a) results. The graph shows the ‘effective’ transmission (and mode conversion) of both tensile and shear waves through the front water/Perspex and rear Perspex/water interfaces for a fixed, small region on the sample.

Similarly, the variation of the time-of-flight with angle of incidence for the same Perspex sample is plotted in Fig. 6.26(b). Note the distinctly different

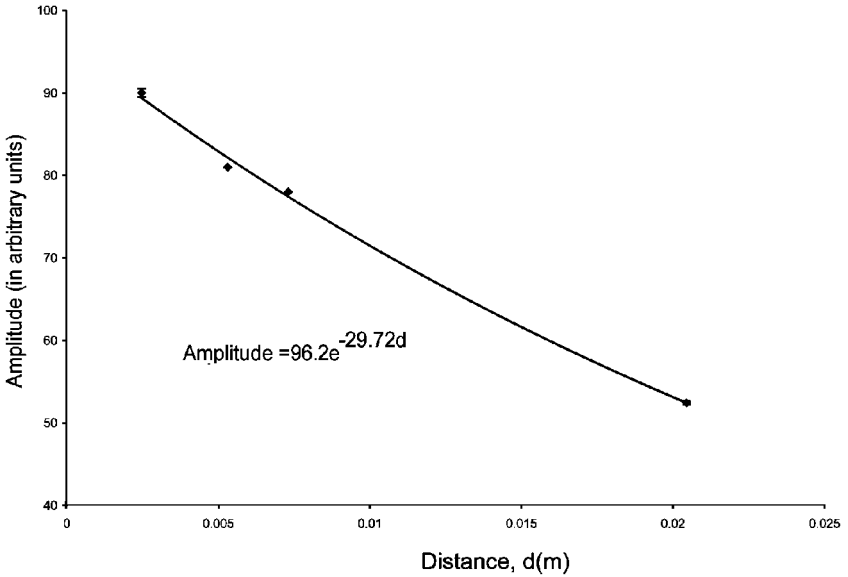


Figure 6.23 The attenuation of ultrasound intensity as a function of the thickness, d, of a parallel-sided Perspex block. The data were taken at an ultrasonic frequency of 2.25 MHz.

regions in both plots. For angles of incidence around 0° (i.e. normal incidence) the times-of-flight are determined by the longitudinal (tensile) waves. For angles of incidence between 38° and 55°, the times-of-flight of the pulse group echoes are determined by the transverse (shear) waves in Perspex.

Using the equations above, from these time-of-flight data, values of the velocities of the longitudinal wave, V_L, and the transverse wave, V_S, as functions of angle of refraction, r, may be computed. The stiffness constants are then found using Christoffel's equations (see below), which link the velocity of longitudinal or tensile waves, V_L, and the velocity of transverse or shear waves, V_S, to four stiffness constants (C_{pq}) for different angles of refraction, r, and sample density, ρ.

$$V_L^2 = \frac{B_1 + B_2 + \sqrt{(B_2 - B_1)^2 + 4B_3^2}}{2\rho} \qquad [6.27]$$

and

$$V_S^2 = \frac{B_1 + B_2 - \sqrt{(B_2 - B_1)^2 + 4B_3^2}}{2\rho} \qquad [6.28]$$

where

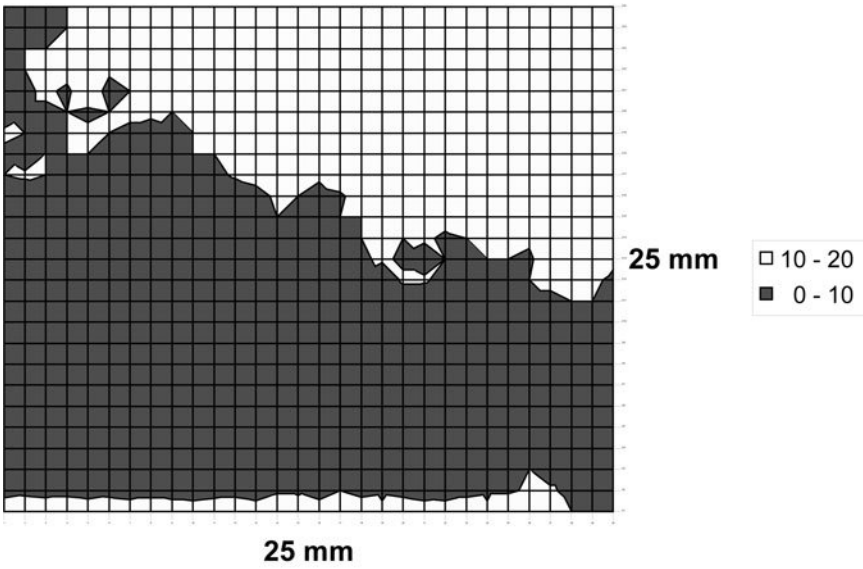


Figure 6.24 Typical TOF area scan for Perspex at normal incidence showing featureless mesostructural effects. The 20 nanosecond TOF variations are most likely due to a systematic variation of thickness across this area of the sample.

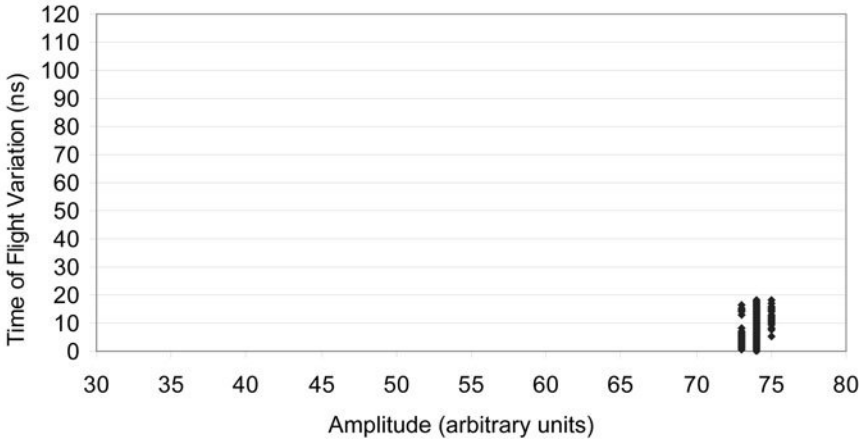


Figure 6.25 By plotting the TOF against amplitude correlation for the same Perspex sample as in Fig. 6.24 (normal incidence), it can be seen that the variations in amplitude are essentially negligible.

$$B_1 = C_{11}\cos^2 r + C_{55}\sin^2 r \quad [6.29]$$

$$B_2 = C_{55}\cos^2 r + C_{33}\sin^2 r \quad [6.30]$$

$$B_3 = (C_{55} + C_{13})\cos r \sin r \quad [6.31]$$

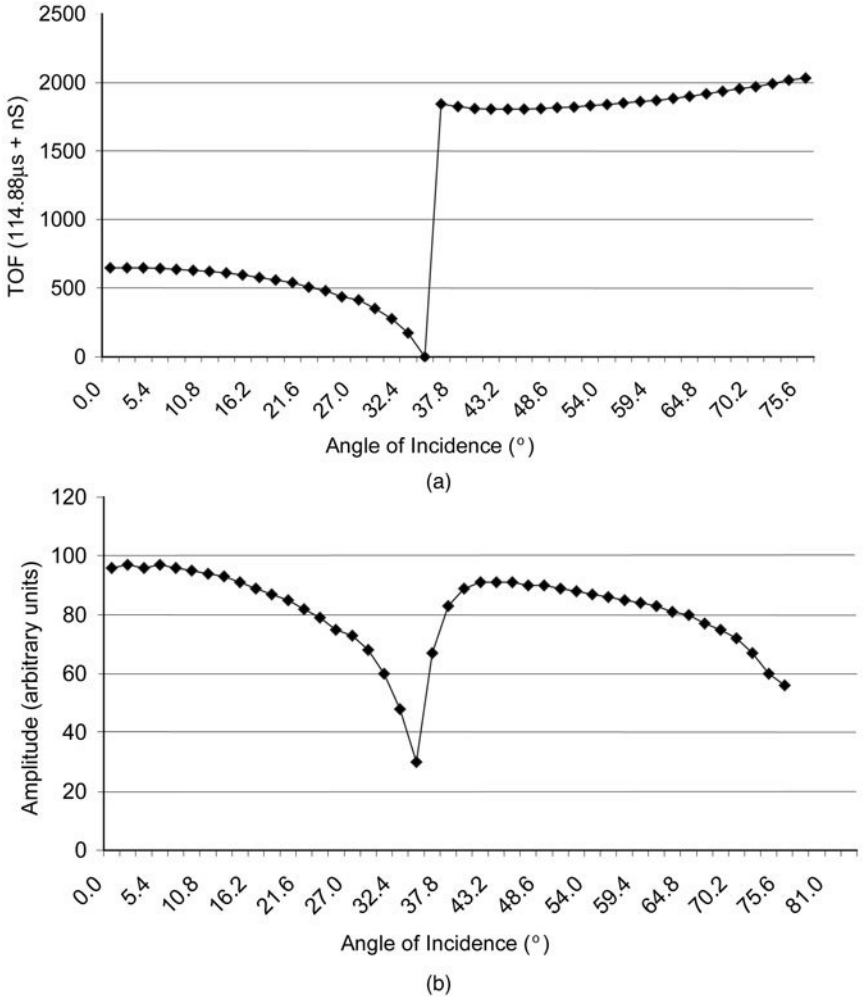


Figure 6.26 The variation of (a) TOF with the angle of incidence and (b) amplitude with angle of incidence for a 3 mm thick Perspex sample is shown. The tensile wave is responsible for the data between 0 and 37.8° and the shear wave is dominant beyond 37.8°.

The stiffness constant subscripts in the equations above assume rotation of the sample about the 2-axis.

Note that there is no unique solution to these equations. A four-dimensional minimisation routine is employed[50] to find acceptable solutions (which minimise some suitable error function) to the measured variation in velocity with angle of refraction. An initial rough guess is made for each stiffness constant's value and, through an iterative process, best values of the stiffness

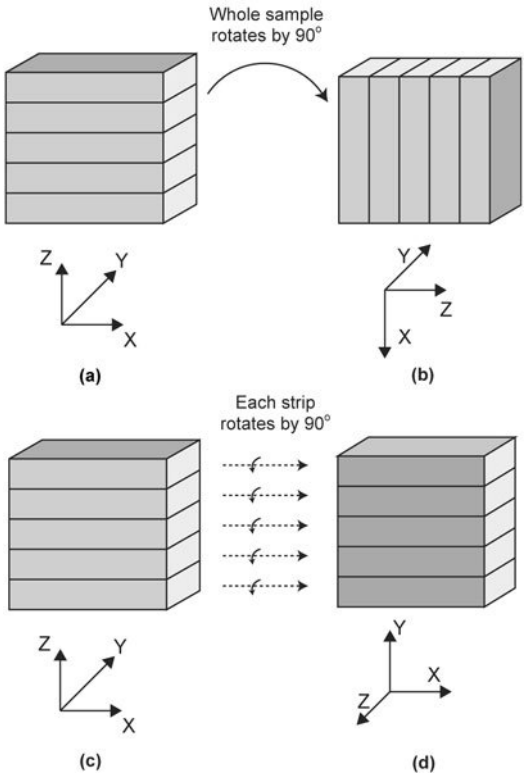


Figure 6.27 A schematic diagram is shown of the measurement of all 3D stiffness constants within a parallel-sided plaque. (a) Initial orientation of the sample, which is scanned over the area at different angles of incidence. (b) After the sample is rotated by 90°, it is rescanned over the area at different angles of incidence. (c) The original orientation is cut into strips as shown and each strip is rotated by 90° before being re-assembled. (d) The re-assembled strips are then scanned again at different angles of incidence.

constants are found. The same idea can be expanded into a five-dimensional minimisation routine, if one wants to explore the effect of small variations in the density value as well as small variations in the stiffness constants.

The other 3D stiffness constants may be found by repeating the measurement process two more times. If the original orientation of the sample is as shown in Fig. 6.27(a), by rotating the sample by 90° (Fig. 6.27(b)) and performing another angular scan, a different set of elastic constants is deduced. Note that if the symmetry of the sample is such that all nine elastic constants need to be obtained, the sample must be cut into strips, the strips rotated and re-assembled, as shown in Fig. 6.27(d), before finally performing another angular scan.

For rotations about the 1-axis, the stiffness constants C_{11}, C_{33}, C_{55} and C_{13} are substituted by C_{22}, C_{33}, C_{44} and C_{23} and, for rotation about the 3-axis, the stiffness constants C_{11}, C_{33}, C_{55} and C_{13} are substituted by C_{22}, C_{11}, C_{66} and C_{12} respectively. The automation of this technique, which was tedious to perform by manual means, decreases the sample acquisition and analysis time by a factor of 15 to 20 and increases the sensitivity of the system. Also, the large area scan capability allows the user to explore the linkage between ultrasonic propagation and sample mesostructure, which would be impracticable with a manual system.

The complete set of Perspex stiffness constants (in GPa), which have been measured by this water immersion technique at 25ºC and an operating frequency of 2.25 MHz are shown below:

$$\begin{pmatrix} 8.16(\pm0.02) & 3.85(\pm0.02) & 3.94(\pm0.01) & & & \\ 3.85(\pm0.02) & 8.57(\pm0.17) & 4.17(\pm0.02) & & & \\ 3.94(\pm0.01) & 4.17(\pm0.02) & 8.44(\pm0.22) & & & \\ & & & 2.25(\pm0.01) & & \\ & & & & 1.99(\pm0.01) & \\ & & & & & 1.99(\pm0.03) \end{pmatrix}$$

The uncertainties in the Perspex stiffness constants are between $\pm0.2\%$ to $\pm1.7\%$ but are difficult to justify unambiguously. They have been calculated from the known percentage timing error and the percentage error in both sample density and sample thickness. Of particular concern, however, is the estimation of the errors introduced by the minimisation process! When heterogeneous samples like fibre-reinforced composites are considered, the situation is potentially more complex, but the area scan approach may be used to indicate the underlying random error of measurement from spatially coherent variations in stiffness due to mesostructural effects, as discussed below.

6.4.4 Measurements on continuous, carbon fibre-reinforced composites

While the Perspex sample shows transverse isotropy, a markedly different sample is a typical unidirectional carbon fibre plaque. The nearly parallel fibres exhibit very high stiffness in the fibre orientation direction and lower stiffness in a direction perpendicular to the fibre orientation. The set of curves which illustrate the relationship between TOF and incident beam angle for different fibre orientation directions is shown in Fig. 6.28.

In order to assess the response of the ultrasonic system to known variations in mesostructure, a unidirectional carbon fibre plaque had a circular hole (of diameter 2 cm) cut out of it and hence a circular insert with variable fibre orientation could be investigated. The TOF area maps, when the fibre orientation in the insert was at 5º to the rest of the plaque, are shown in Fig. 6.29(a) for tensile waves at normal incidence and in Fig. 6.29(b) for shear wave propagation

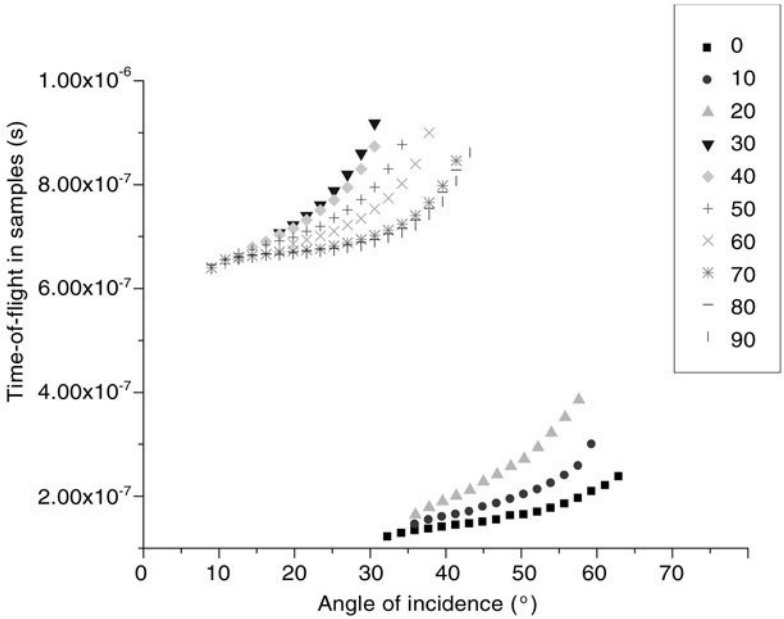


Figure 6.28 The shear wave TOF variations at a fixed point on a unidirectional, carbon fibre-reinforced sample are shown. As the orientation of the carbon fibres is rotated in 10° steps from 0 to 90°, there is an observed, dramatic variation in both TOF (and also amplitude) as the angle of the incident beam is altered.

(at an incident angle of 40°). As expected, the contours in these figures show a roughly circular shape for normal incidence and an elliptical shape for the shear wave propagation. The gradients seen in line scans through the centre of the insert (Fig. 6.30(a) and Fig. 6.30(b)) may be explained by the effect of the 1 cm diameter ultrasonic beam gradually overlapping the differing fibre orientations in plaque and insert.

In short fibre composites, the situation would be more complex due to the range of fibre orientations within the sample. An experiment was performed where a thin unidirectional carbon fibre plaque was placed at 35° to the ultrasonic beam (to stimulate the shear waves) and the fibre orientation in the plaque was rotated every 10° before taking another linescan. In this way, data were obtained on the change in TOF and amplitude with varying fibre orientation and these data are plotted as a correlation graph in Fig. 6.31. In effect, this plot represents a limit to the range of TOF and amplitude that might be expected from a similar thickness but more complex sample in which different carbon fibre orientation states are present. These observed variations in TOF and amplitude are far greater than variations due to systematic mould thickness differences of the order of many tens of microns across the sample.

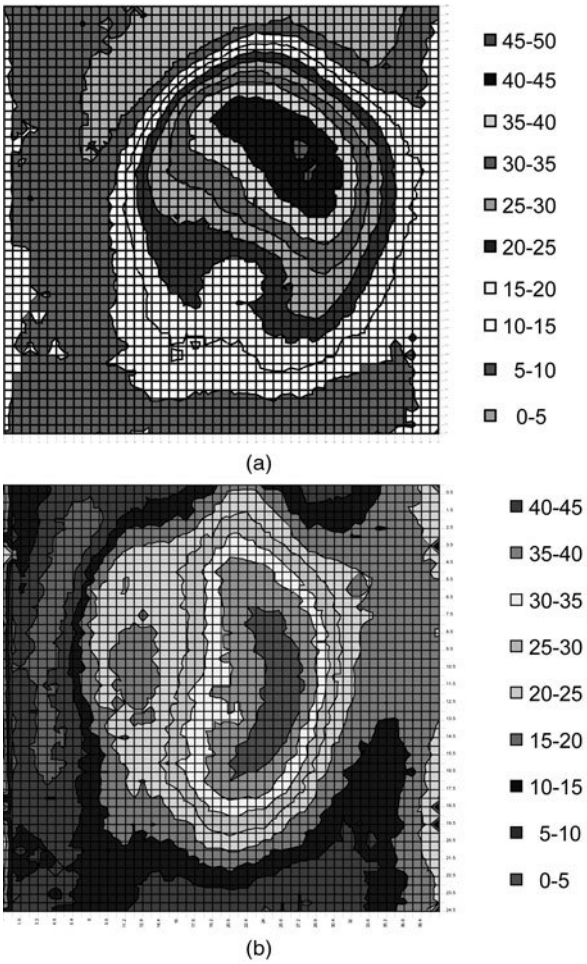


Figure 6.29 (a) TOF map due to tensile waves at normal incidence propagating through a thin unidirectional carbon fibre plaque (which has a 2 cm diameter insert). (b) The TOF map for the same sample due to shear waves propagating at 40°. The carbon fibres in the insert were oriented at 5° to those fibres in the rest of the plaque. Both scans covered an area of 2.5 cm × 2.5 cm and the time of flight (TOF) is given by the number of nanoseconds from an offset of 115.7 μsec.

6.4.5 Short glass fibre-reinforced, injection-moulded composites

Our main objective, over the past few years, has been to acquire the stiffness constant maps for more complex composites like injection-moulded glass fibre-reinforced polymers. A number of samples, processed under different conditions by colleagues in Dr Fischer's Group at IKP (Stuttgart) and latterly by Dr Vincent's Group at CEMF (Sophia-Antipolis), have been studied with the Leeds

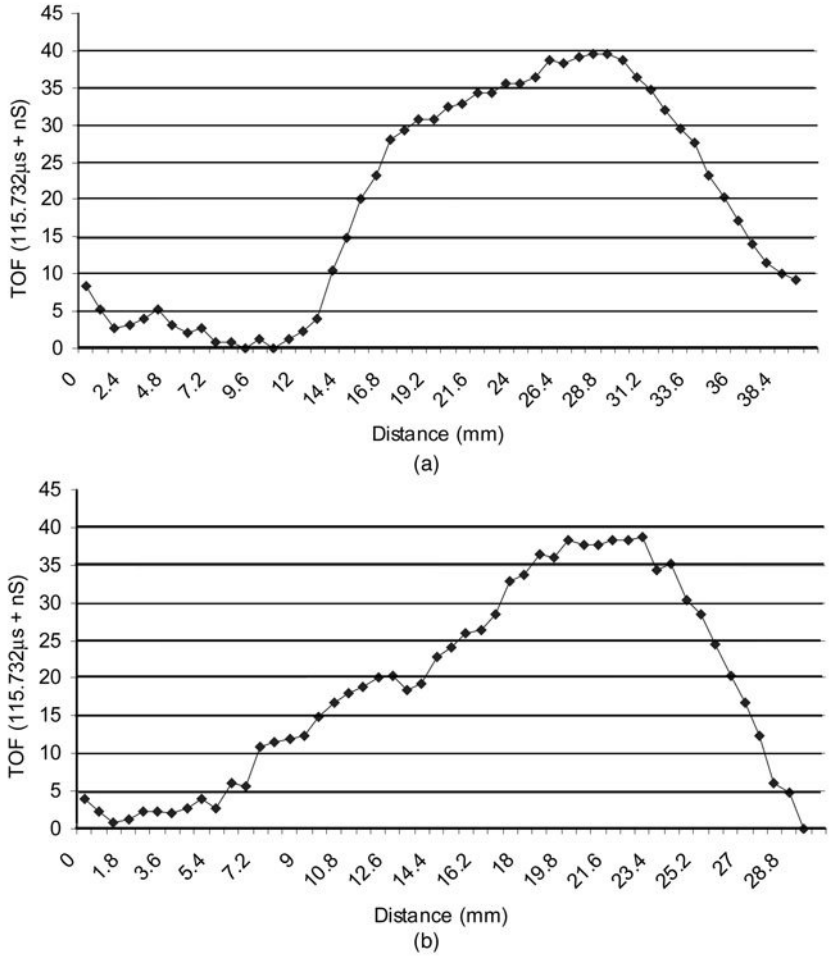


Figure 6.30 Perpendicular line scans of the TOF across the circular carbon fibre insert at normal incidence (Fig. 6.29(a)) to illustrate the response of the system when scanning over two different regions of fibre orientation. The TOF gradients at the edges of the insert are due to the overlap of the 1 cm diameter beam width with the 2 cm diameter insert.

ultrasonic water immersion system. The project aims were two-fold: to compare the actual stiffness constants with the predictions made by software modelling packages like Moldflow™, and to explore the reproducibility of mechanical properties (by acquiring the stiffness constants) from one sample to the next.

Figure 6.32(a) shows a schematic of a typical, 5 mm thick plaque sample created by a multiple, push-pull injection-moulding process and a 200 μm $\times$ 5 mm region scanned by the 2D image analyser (discussed in detail in Chapter 3). Using the 2D system, the fibre orientations along a cross-section

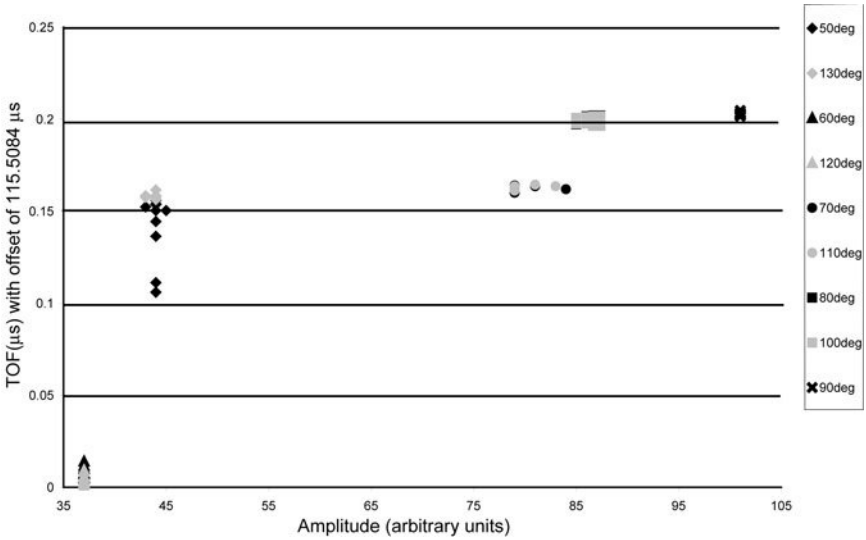


Figure 6.31 Correlated TOF and amplitude data are shown for shear waves at an incident beam angle of 35°. The data were obtained by varying the orientation of the fibres in a thin, unidirectional, carbon fibre plaque from 50° through to 130° in 10° steps. The sensitivity of both shear wave attenuation and propagation delays to mesostructural variations in the fibre orientation state is clear. These data imply that the different, localised, fibre orientation states observed in short carbon fibre composite samples would also give similar correlation plots.

may be deduced – as shown in Fig. 6.32(b), where the orientation tensor coefficients are plotted. As shown in this figure, these samples have a skin-core-skin cross-sectional microstructure, but, depending on the process parameters and the thickness of the moulds employed, there may be significant variations in the thickness of the core regions. These core thickness variations and attendant fibre orientation differences should influence the propagation of the ultrasound shear waves, as indicated by the carbon fibre insert test. In effect, the mesostructural differences in the fibre orientation state will create spatial variations in the stiffness constants.

As an illustration of typical shear wave area scans, consider Fig. 6.33(a–b), (c–d) and (e–f), which shows TOF data for three regions: (a–b) near an injection gate, (c–d) in the middle of the plaque and (e–f) further away from the injection gate. The two glass fibre-reinforced polypropylene samples were processed during the same run. Morphologically, there are similarities and also differences over these $2\,\text{cm} \times 2\,\text{cm}$ areas. The TOF and amplitude for each of these three regions on one of the two samples have been superimposed in the correlation plot of Fig. 6.34. Note the gradual shifting of the data but the retention of the characteristics of the overall correlations. The shifting could be due to gradual changes in mould thickness or warpage effects and the dramatic variations in both TOF and amplitude bears a resemblance to the carbon fibre data in Fig.

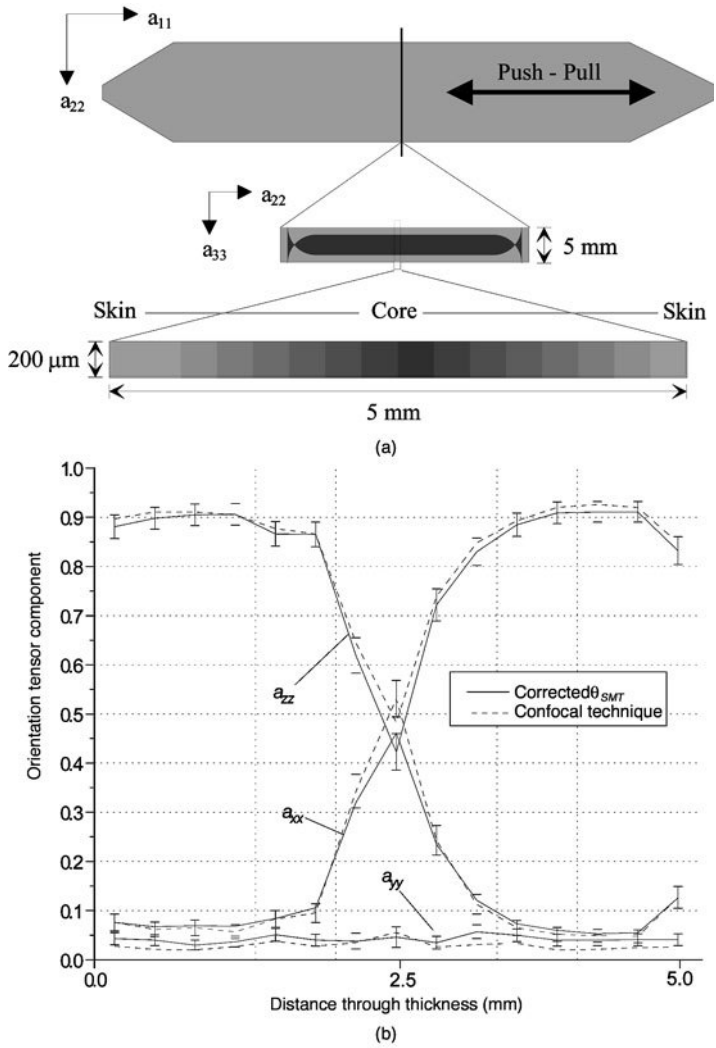


Figure 6.32 (a) A schematic of an injection-moulded, glass fibre-reinforced plaque is shown which exhibits a skin-core-skin structure. (b) By sectioning and analysing the fibre orientations with the Leeds 2D image analyser, these skin-core-skin regions are clearly illustrated by the variation in fibre orientation tensor coefficients across the sample thickness.

6.31. This implies that local fibre orientation states and possibly variations in core thickness are causing these variations. Note that these area scans may be used to investigate reproducibility between samples processed in an identical way (and the authors are not aware of any systematic study having been carried out to look at the reproducibility of stiffness in fibre-reinforced composite parts).

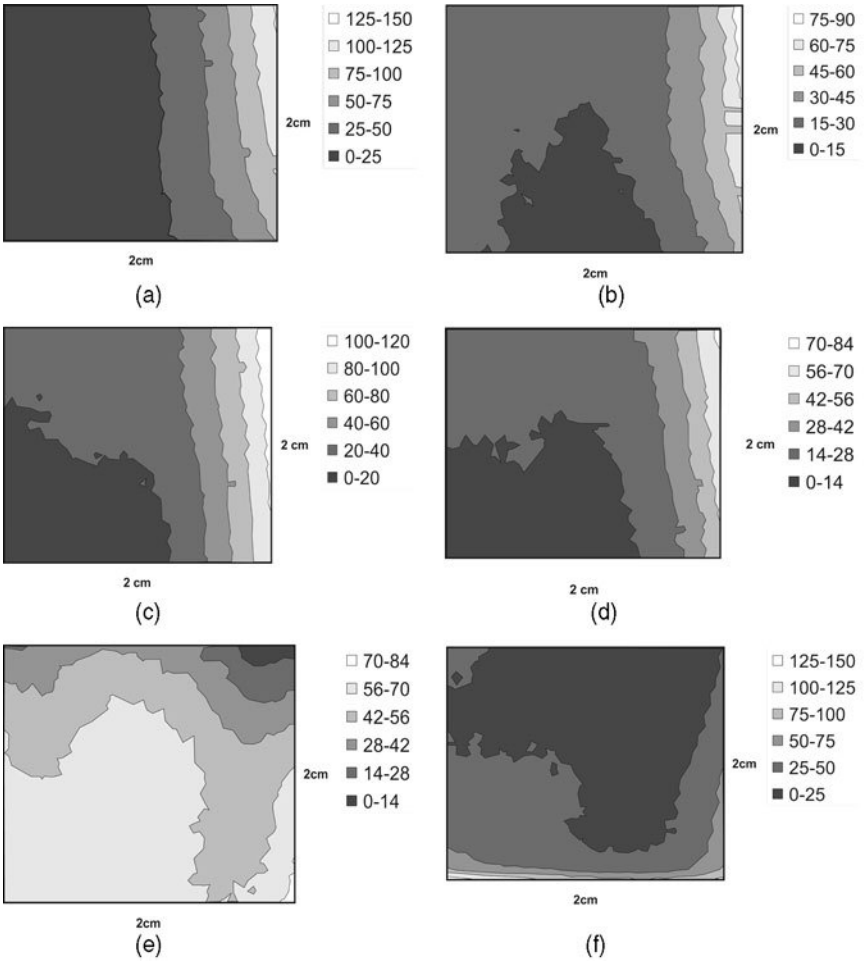


Figure 6.33 The variation of the TOF at different positions on injection-moulded, glass fibre reinforced polypropylene plaques may be investigated (a) and (b) close to an injection gate, (c) and (d) in the middle of the plaque and (e) and (f) further away from the injection gate. The same regions on two identically processed samples are shown for an incident beam angle of 35° and the TOF is given as a nanosecond delay from an offset of 119.96 μsec.

By generating area scans at different angles of incidence (in the range from 0° through to 60°), the data may be used to produce area maps of each stiffness constant. Examples of two such maps are shown in Fig. 6.35(a) and (b). Clearly, there appear to be regions within each map where the stiffness constant is higher than other parts of the area map. In Fig. 6.36(a) and (b), the corresponding frequency distributions of stiffness constant values for the two area maps are shown. These plots show that the random measurement errors are much smaller than the variations due to material structure – which would not become apparent if a few (manual) spot

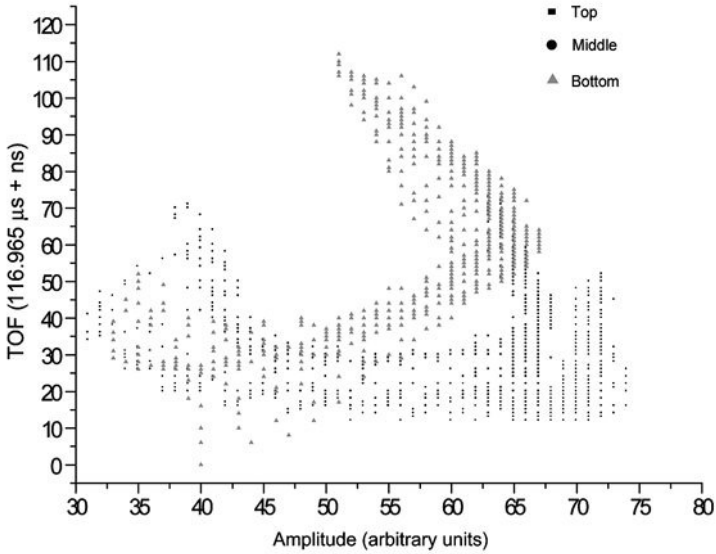


Figure 6.34 Correlation plots of TOF against amplitude for the three regions – top, middle and bottom – on one of the samples shown in Fig. 6.33. The relative shift between the three plots (from high amplitude and low TOF to lower amplitude and higher TOF) is probably due to variations in mould thickness but the overall shape of the correlations indicate the local variations in glass fibre orientation states, cf Fig. 6.31.

measurements were to be taken at random within these areas. For example, the mean value of C_{XX} over this area of the sample is 19.5 GPa (with the random measurement error being less than $\pm 2\%$) but it exhibits a mesostructural variability of $\pm 20\%$.

6.4.6 Conclusions

Ultrasonic measurements have been carried out for decades and it might be thought that the research field was fully understood with no surprises in store. However, for heterogeneous materials like fibre-reinforced polymer composites many issues have to be addressed. For example, there are many different formulations of the predictive equations for reflection and transmission coefficients at the liquid/composite and the composite/liquid interfaces. At least four different formulations for the propagation across the liquid/solid interface are known to the authors; see Shutilov,[51] Mayer,[52] Krautkramer[53] and, more recently, Nayfeh.[42] We have yet to reconcile completely any of these predictive equations with the observed, typical variation in beam amplitude as a function of angle of incidence, as shown in the previous Perspex sample (Fig. 6.26(b)).

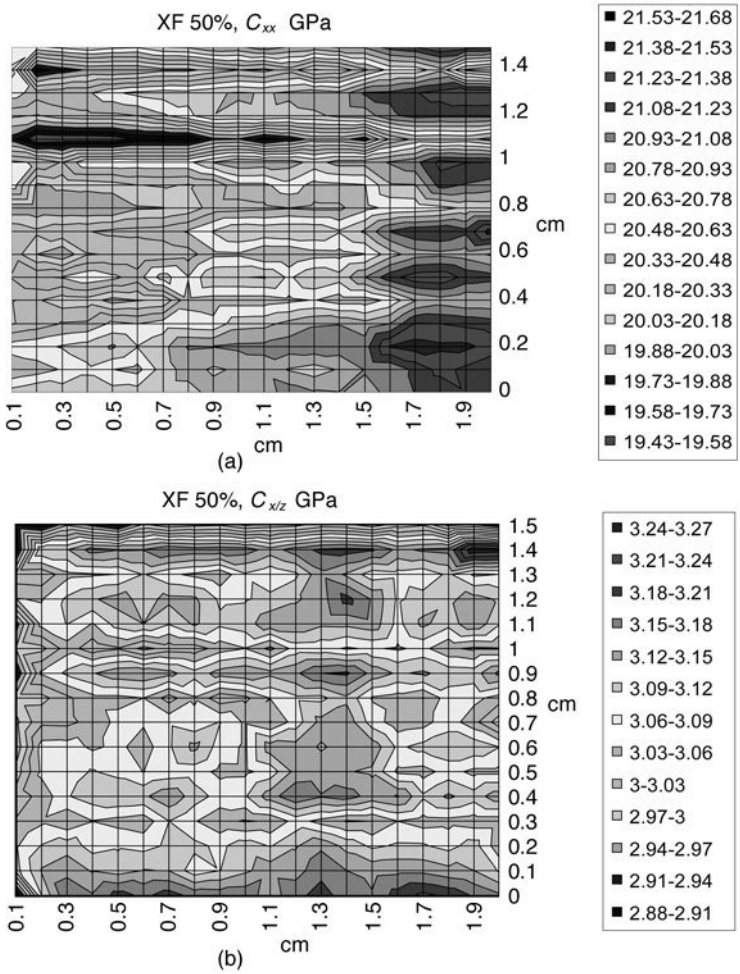


Figure 6.35 Stiffness constant maps may be derived from a number of TOF area scans taken at different angles of incidence. Some typical maps are shown for a glass fibre-reinforced polyarylamide, injection-moulded sample. The stiffness constants, (a) C_{XX} and (b) $C_{X/Z}$ illustrate mesostructural variations. All stiffness values are in GPa.

At the moment, there are suggestive correlations between fibre orientation states and ultrasonic wave propagation, but more investigations are needed and more research into exactly how one could use these low frequency signals to work backwards and infer the fibre orientation states within the sample with reasonable confidence.

Even though there is not enough information yet to link fibre orientation state to ultrasonic propagation in these complex parts, this is an excellent technique for exploring the sample-to-sample variations in the common manufacturing

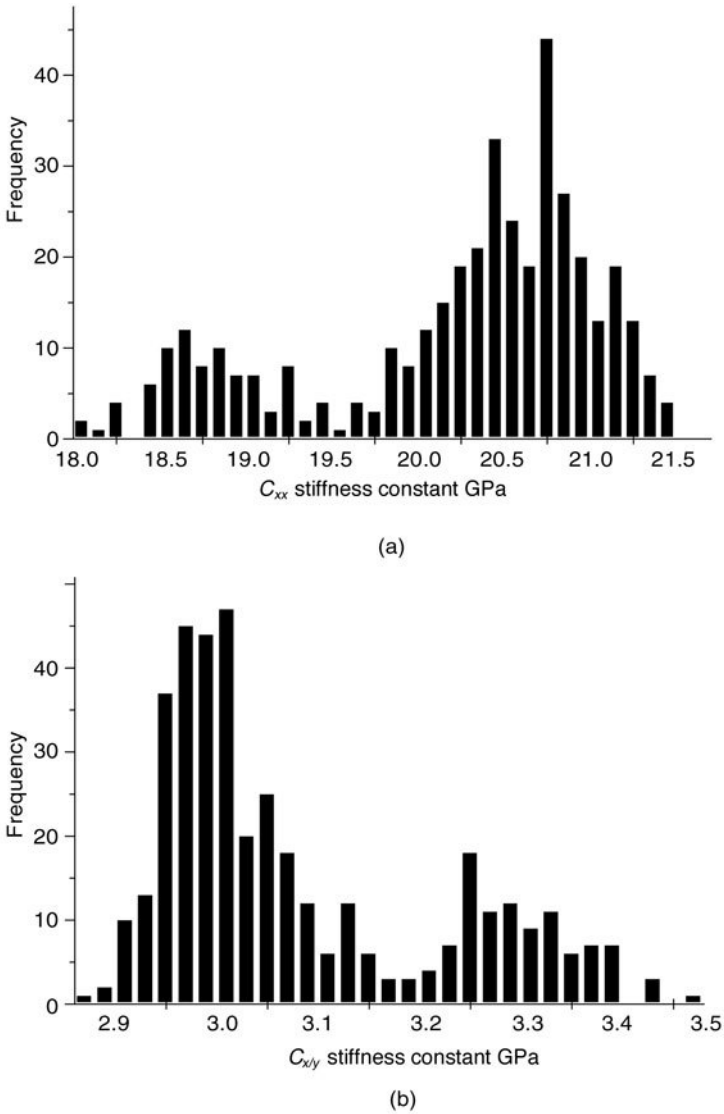


Figure 6.36 Frequency distributions of the stiffness constant values (a) C_{XX} and (b) $C_{X/Z}$, from Fig. 6.35, showing that the random measurement error for a particular region is much smaller than the variations due to mesostructural effects over that sample area.

processes like injection moulding. By automating this water immersion testrig, there has been a significant speed-up of the operation and hence it is far easier

- to establish the range of stiffness constants for each component part
- to detect any important sub-surface features

- to differentiate systematic mould thickness effects from the underlying manufacturing process spatial variations, as shown in stiffness constant maps.

Finite element modelling packages are still at a crude state of development in terms of 3D simulations. The actual 3D stiffness data from ultrasonic systems, such as the Leeds testrig, will serve both as a benchmark for simulation predictions and also help improve/justify the assumptions made in these software packages.

We have shown that, by scanning the samples at a fixed angle of incidence ($\theta_i = 35^o$), the propagation of the ultrasonic shear waves is sensitive to mesostructural variations through the sample thickness. Hence, the area scanning mode may be used to identify interesting regions for further study of fibre orientation state with the 2D image analyser or confocal laser scanning microscope. Similarly, by scanning the samples at a range of angles of incidence, the stiffness constants may be deduced and hence better estimates of sample stiffness can be obtained, as well as more representative errors.

Research is continuing to find a model for the propagation of these low frequency ultrasonic waves in industrially important, heterogeneous materials like composites. It is believed that ultrasonic propagation through injection-moulded, fibre-reinforced composites can be understood in terms of a modified, simple layer model which takes into account the fibre orientation states.[42]

6.5 Epilogue

There are so many techniques, some of which achieve spatial resolutions equal to, or better than, optical microscope resolutions and there are other techniques, which enable the scanning over significant sample volumes. Unfortunately, these techniques are outside the scope of this book, but interested readers may wish to look into them further. For example, in the recent literature, the following techniques have been mentioned:

- *Secondary ion mass spectrometry (SIMS)*, Chandra[54]
- *Eddy current detection*, Zorgati *et al.*[55]
- *Capacitative array sensing*, Shull *et al.*[56]
- *Positron annihilation spectroscopy*, Park *et al.*[57]
- *3D ion microtomography*, Sakellariou *et al.*[58]
- *3D atom probe technique*, Vurpillot *et al.*[59]
- *Neutron radiography*, Mast *et al.*[60]

This research area – the microscopy of materials – has expanded rapidly over the past few decades as more sensitive detectors became available and the computing power of the most economical personal computers (or workstations) has increased. It is difficult to keep up to date with developments, but it is hoped that this book has given an indication of the vitality in this subject. However,

there are many excellent academic journals, which ensure that the latest developments are reported in depth, including the *Journal of Microscopy*, *Scanning*, *Reviews of Modern Physics* and *Reviews of Scientific Instrumentation*.

This book has been concerned with (mainly optical) computer-aided microscopy measurement techniques for materials science. An interesting recent development is the idea of web-based 'Virtual Instruments'. In Chapter 2, mention was made of the Vela laboratory aid, which incorporated a Motorola microprocessor – the M6802 – which, with a little hardware, mimicked the performance of 'real instruments' that were found in the typical school science laboratory (and a few university laboratories too)! The simple idea was that any real instrument could be simulated in software (even if the operational limitations were inferior to the real thing). At the same time as the Vela development, National Instruments Inc. was developing a 'Virtual Instrument' concept on personal computers. In this scheme, by plugging in a special data acquisition and control card and running National Instruments software, virtual instruments could be manipulated and interconnected on screen, before making the appropriate measurements on a real experiment.

With the emergence of the Internet and inexpensive global networking, other interesting possibilities present themselves in connection with virtual instruments for the user:

- *Visualisation of images at a distance* – video conferencing from different parts of the world or access to microscope images at any time for analysis
- *Remote control of operations* – there are a growing number of microscopes which are controllable over the Internet. For example, Tanaka *et al.*[61] have remote control of their electron diffraction instrument.
- *Opportunities for T&L* – experiments involving scheduling of time on instruments for schoolchildren, industrialists or researchers.

But what about the near future – how could the current developments in artificial intelligence, robotics, quantum computing, miniature light sources, global information networks and smart materials influence the direction of microscopy? How will miniaturisation into the nanometre scale size regime affect the techniques that have been mentioned in this book? It is interesting to speculate what might transpire in the next few decades.

So to end, we would like to present a few plausible scenarios for the future of microscopy!

Scenario 1: Future merging of the computers and instrumentation will probably enable the researcher to take out of his/her pocket a foldable, polymeric-based flexible sheet. This sheet would be a display/control panel with minimal computer power but which has a radio telemetry link to a home-based, or work-based, server. The polymer portal could give the user access to their laboratory-

based microscope system and allow the automated control of sample selection, microscope control and image acquisition and initiate the image analysis from anywhere in the world.

Scenario 2: Alternatively, perhaps the researcher will visit an industrial client and be able to take out of his/her pocket a palm-size device, which is a nano-engineered, optical microscope. The miniature microscope is placed on the specimen to be analysed and it will move over the surface of that specimen, taking 3D confocal images at different depths. Other possible mutants for nano-engineered microscopes could include a suite of AFM designs. Is this a far-fetched idea? Remember the reference in Chapter 4 for the smallest confocal microscope design!

But, no doubt, the future for microscopy in the twenty-first century will be even more interesting and exciting than these scenarios suggest. Microscopy has always had new surprises in store and has always been capable of incorporating the latest technological discoveries. The ultimate goal of '... reliable reconstruction of materials and different chemical species in 3D in order to infer the mesostructural regions within samples' is rapidly becoming a reality – then, no doubt, the overriding aim will be to perform these 3D reconstructions even faster, more efficiently and, of course, even more cheaply.

6.6 References

1 Klemperer O. (1961), *Electron Physics: The physics of the free electron*, Butterworths, London.
2 Everhart T.E. and Hayes T.L. (1972), The Scanning Electron Microscope, *Scientific American*, January, 54.
3 Grim P.C.M. and Hadziioannou G. (1994), Imaging and characterisation of materials by the new scanning probe techniques (STM/AFM), in *Characterisation of Polymer Composites*, ed. H. Ishida, Butterworth-Heinemann, Chapter 7, pp. 129–146.
4 Binnig G., Rohrer H., Gerber C.H. and Weibel E. (1982), Surface studies by scanning tunneling microscopy, *Phys. Rev. Lett.*, **49**, 57–61.
5 Binnig G., Quate C.F. and Gerber C.H. (1986), Atomic force microscopy, *Phys. Rev. Lett.*, **56**, 930–933.
6 Reimer L. (1989) *Transmission Electron Microscopy*, Springer-Verlag, Berlin.
7 Den Dekker A.J., Sijbers J. and Van Dyck D. (1999), How to optimize the design of a quantitative HREM experiment so as to attain the highest precision, *J. Microsc-Oxford*, **194**, Part 1, 95–104.
8 Amelinckx S., Van Dyck D., Van Landuyt J. and Van Tendeloo G. (1997), *Handbook of Microscopy, Methods I and II*, VCH, New York.
9 Danilatos G.D. (1993), Introduction to the ESEM instrument, *Microsc. Res. Techn*, **25**, 354–361.
10 Howie A. (1995), Recent developments in secondary electron imaging, *J. Microsc-Oxford*, **180**, 192–203.
11 Thiel B.L., Bache I.C., Fletcher A.L., Meredith P. and Donald A.M. (1997), An

improved model for gaseous amplification in the ESEM, *J. Microsc-Oxford*, **187**, 143–157.
12 Stelmashenko N.A., Craven J.P., Donald A.M., Terentjev E.M. and Thiel B.L. (2001), Topographic contrast of partially wetting water droplets in environmental scanning electron microscopy, *J. Microsc-Oxford*, **204**, Part 2, 172–183.
13 www.leica-microsystems.com, www.oxford-instruments.com and for historical information on electron microscopy, see www-g.eng.cam.ac.uk/125/achievements/semhist-intro.htm
14 Dillard J.G. (1994), X-Ray Photoelectron Spectroscopy (XPS) and Electron Spectroscopy for Chemical Analysis (ESCA), in *Characterisation of Composite Materials*, ed. H. Ishida, Butterworth-Heinemann, Chapter 1.
15 Paparazzo E. (1994), Applications of scanning Auger microscopy to the analysis of materials, *Microscopy & Analysis*, Nov, 17–19.
16 Ferguson I.F. (1989), *Auger Microprobe Analysis*, Adam Hilger, Bristol.
17 Vajda E.G. (1998), Errors in quantitative back-scattered electron analysis of bone standardised by energy-dispersive X-ray spectrometry, *Scanning*, **20**, 527–535.
18 Farhang H., Napchan E. and Blott B.H. (1993), Electron back-scattering and secondary electron emission from carbon targets: comparison of experimental results with Monte-Carlo simulations, *J. Phys. D: Appl. Phys.*, **26**(12), 2266–2271.
19 Yasuda M., Kawata H. and Murata K. (1996), The spatial distribution of back-scattered electrons calculated by a simple model, *Phys. Stat. Sol. (a)*, **153**(1), 133–144.
20 Napchan E. (2001), Back-scattered electrons in the SEM, *Microscopy & Analysis*, January, 9–11.
21 Terauchi M., Tanaka M., Tsuno K. and Ishida M. (1999), Development of a high energy resolution electron energy-loss spectroscopy microscope, *J. Microsc-Oxford*, **194**, Part 1, 203–209.
22 Leapman R. (1992), *EELS Quantitative Analysis Transmission Electron Energy Loss Spectrometry in Materials Science* (eds. M.D. Disko, C.C. Ahn and B. Fulz) pp. 47–83, Electronic, Magnetic and Photonic Materials Division monograph series no. 2, The Minerals, Metals and Materials Society, Warrendale.
23 Lakner H., Rafferty B. and Brockt G. (1999), Electronic structure analysis of (In, Ga, Al) N heterostructures on the nanometre scale using EELS, *J. Microsc-Oxford*, **194**, Part 1, 79–83.
24 Liu C.P., Preston A.R., Boothroyd C.B. and Humphreys C.J. (1999), Quantitative analysis of ultrathin doping layers in semiconductors using high-angle annular dark field images, *J. Microsc-Oxford*, **194**, Part 1, 171–182.
25 De Graef M., Nahfer N.T. and McCartney M.R. (1999), Phase contrast of spherical magnetic particles, *J. Microsc-Oxford*, **194**, Part 1, 84–94.
26 Knippelmeyer R. and Kohl H. (1999), Relativistic calculations of intensity distributions in elemental maps using contrast transfer functions, *J. Microsc-Oxford*, **194**, Part 1, 30–41.
27 Zipprich J., Fuller T., Banhart F., Schmidt O.G. and Eberl K. (1999), The quantitative characterisation of SiGe layers by analysing rocking profiles in CBED patterns, *J. Microsc-Oxford*, **194**, Part 1, 12–20.
28 Yakimov E.B. (2001), Modulated EBIC and photocurrent: diffusion length depth distributions, *Microscopy & Analysis*, January, 21–23.
29 Callaghan P.T. (1991), *Principles of Nuclear Magnetic Resonance Microscopy*, Oxford: Clarendon Press.

30 Koenig J.L. (1992), *Spectroscopy of Polymers*, ACS, Washington, Chapter 11.
31 Matzkanin G.A. (1989), A review of nondestructive characterisation of composites using NMR, in *Nondestructive Characterisation of Materials*, eds P. Höller, V. Hauk, G. Dobmann, C. Ruud and R.E. Green, Springer-Verlag, New York, pp. 655–669.
32 Blumich B. and Blumler P. (1993), NMR imaging of polymer composites, *Makromol. Chem.*, **194**, 2133–2161.
33 Hoh K-P., Perry B., Rotter G., Ishida H. and Koenig J. (1989), Anlaysis of composite processing using magnetic resonance imaging, *J. Adhesion*, **27**(4), 245–249.
34 Jezzard P., Wiggins C., Carpenter T., Hall L., Barnes J., Jackson P. and Clayden N. (1992), Demonstration of NMRI for void detection in carbon-fibre reinforced polymer composites and comparison with ultrasound methods, *J. Mater. Sci*, **27**(23), 6365–6370.
35 Briggs A. (1992), *Acoustic Microscopy*, Monographs on the Physics and Chemistry of Materials, Clarendon Press, Oxford.
36 Li W. and Hoyle B.S. (1991), Ultrasonic process tomography using multiple active sensors for maximum real-time performance, *Chemical Engineering Science*, **52**(13), 2161–2170.
37 Rabe U., Kopyinska M., Hirsekorn S. and Arnold W. (2001), Evaluation of the contact resonance frequencies in Atomic Force Microscopy as a method for surface characterisation, *Proc Ultrasonics International '01, Delft July 2001*, Paper B1.01.
38 Dunn F. and Fry W.J. (1959), Ultrasonic absorption microscope, *J. Acoustic Soc. Amer.*, **31**, 632–633.
39 Lisy F., Hiltner A., Baer E., Katz J.L. and Meunier A. (1994), Application of scanning acoustic microscopy to polymeric materials, *J. App. Polymer Sci.*, **52**, 329–352.
40 Fatkin D.G.P., Scruby C.B. and Briggs G.A.D. (1989), Review: acoustic microscopy of low-ductility materials, *J. Mat. Sci.*, **24**, 23–40.
41 de Belleval J.F., Boyer Y. and Lecupu D. (1989), Porosity characterisation in thin composite plates by ultrasonic measurements, in *Nondestructive Characterisation of Materials*, eds P. Höller, V. Hauk, G. Dobmann, C. Ruud and R.E. Green, Springer-Verlag, New York, pp. 131–139.
42 Nayfeh A.H. (1995), *Wave Propagation in Layered Anisotropic Media*, North Holland Series in Applied Mathematics and Mechanics, Vol. 39, Elsevier Science B.V.
43 Ward I.M. (1985), *Mechanical Properties of Solid Polymers*, 2nd edn, John Wiley & Sons.
44 Markham M.F. (1970), Measurement of the elastic constants of fibre composites by ultrasonics, *Composites*, March, 145–149.
45 Read B.E. and Dean G.D. (1978), *The Determination of the Dynamic Properties of Polymers and Composites*, Adam Hilger, chapter 7, pp. 162–179.
46 Lord D. (1989), The determination of the elastic constants of fibre reinforced composites by an ultrasonic method, PhD Thesis, Department of Physics, University of Leeds.
47 Hine P.J., Davidson N.C., Duckett R.A., Clarke A.R. and Ward I.M. (1995), Hydrostatically extruded, glass fibre reinforced polyoxymethylene: I The development of fibre and matrix orientation, *Polym. Composites*, **17**(5), 720–729.
48 Enderby M., Clarke A.R., Patel M., Ogden P. and Johnson A.A. (1998), An

automated, ultrasonic immersion technique for the determination of 3D elastic constants of polymer composites, *Ultrasonics*, **36**, 245–249.
49 Harper M.J. and Clarke A.R. (2002), Low frequency ultrasonic propagation through fibre-reinforced, polymer composites, *Ultrasonics*, **40**(1–8) 555–560.
50 Press W., Teukolsky S., Vetterling W. and Flannery B. (1992), *Numerical Recipes in C*, Cambridge University Press, Cambridge, pp. 125–139, 394–455.
51 Shutilov V.A. (1988), *Fundamental Physics of Ultrasound*, Gordon & Breach Science Publishers.
52 Mayer W.G. (1965), Energy partition of ultrasonic waves at flat boundaries, *Ultrasonics*, **3**, 62–68.
53 Krautkramer J., Krautkramer H. and Grabendörfer W. (1990), *Ultrasonic Testing of Materials*, Springer-Verlag, pp. 564–565.
54 Chandra S. (2001), Studies of cell division (mitosis and cytokinesis) by dynamic SIMS ion microscopy: LLC-PK_1 epithelial cells as a model for subcellular isotopic imaging, *J. Microsc-Oxford*, **204**, Part 2, 150–165.
55 Zorgati R., Bernard A., Pons F., Duchene B., Lesselier D. and Tabbara W. (1989), Modelling of the electromagnetic field diffracted by an inhomogeneity in metal: a first step in magnetic imaging, in *Nondestructive Characterisation of Materials*, eds P. Höller, V. Hauk, G. Dobmann, C. Ruud and R.E. Green, Springer-Verlag, New York, pp. 590–597.
56 Shull P.J., Clark A.V. and Auld B.A. (1989), Applications of capacitative array sensors to nondestructive evaluation, in *Nondestructive Characterisation of Materials*, eds P. Höller, V. Hauk, G. Dobmann, C. Ruud and R.E. Green, Springer-Verlag, New York, pp. 582–589.
57 Park Y.K., Lee J.O. and Lee S. (1989), Nondestructive characterisation of a deformed steel using positron annihilation, in *Nondestructive Characterisation of Materials*, eds P. Höller, V. Hauk, G. Dobmann, C. Ruud and R.E. Green, Springer-Verlag, New York, pp. 839–845.
58 Sakellariou A., Jameson D.N. and Legge G.J.F. (2001), Three-dimensional ion microtomography, *Nucl. Inst. & Methods Phys. Res. Sect. B-Beam Interact. Mater.*, **181**, 211–218.
59 Vurpillot F., da Costa G., Menand A. and Blavette D. (2001), Structural analyses in three dimensional atom probe: a Fourier transform approach, *J. Microsc-Oxford*, **203**, Part 3, 295–302.
60 Mast H-U., Brandler T., Knorr E. and Stein P. (1989), Small neutron radiography systems and their applications, in *Nondestructive Characterisation of Materials*, eds P. Höller, V. Hauk, G. Dobmann, C. Ruud and R.E. Green, Springer-Verlag, New York, pp. 707–714.
61 Tanaka M., Tsuda K., Terauchi M., Tsuno K., Kaneyama T., Honda T. and Ishida M. (1999), A new 200 kV Ω filter electron microscope, *J. Microsc-Oxford*, **194**, Part 1, 219–227.

Index

Index terms	Links
A	
A100 digital signal processor	106
Abbé, Ernst	54
aberration function	362
aberrations	58
chromatic	59 231
spherical	59 243 362
absorption coefficient	38
acceptance cone	
objective lens	57
optical fibre	17
accessory plate	73
accommodation of the eye	48
accumulator	100 150
achromatic doublet lens	58
acid digestion	213
acousto-optical deflector (AOD)	139 238 249
adaptive control	236
adaptive neural networks	81
address bus	98
adjustable iris	242
Airy diffraction pattern	37 52 233 262
algorithm	
blending	259
fibre location	347
flood-fill	274

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
algorithm *(Continued)*				
iterative fibre search	286			
recombination	274			
Alhacen	49			
aliasing	127	252		
alphanumeric codes	89			
ambiguity				
in-plane angle	171	189	201	298
amplitude				
fibre	291			
wave	11			
analogue to digital converter (ADC)	99	116		
busy output from	117			
command convert input to	116			
flash-	117			
slow-scan	119			
angular dependence of reflectivity	27			
angular frequency	12	17	373	
anti-blocking agents (silicas)	267			
anti-reflection coating	113			
anti-Stokes Raman signal	308			
aperture function	78			
apertureless SNOM	320			
apertures				
confocal	230	237	241	
diffracting	36	325		
optical probe	319			
apparent depth	243	254		
application specific chips (ASICs)	103			
arclength	292			

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
artefacts			
orientation	174		
polishing	257		
satellite	245	274	
X-ray	341		
ASCII characters	90		
aspect ratio			
fibre	178	213	
image pixel	173		
asperities (bumps)	266		
assembly language	102	131	
astigmatism	231		
atom probe	404		
atomic structure	40	367	
attenuation	38	248	
optical fibre	17		
ultrasonic	390		
X-ray	331	339	
Auger electron spectroscopy (AES)	365		
Auger process	366		
autofluorescence	45	246	255
auto-focus	249	265	
axial range	331		
axial resolution	234	241	
axial response function	236	249	

B

Babbage, Charles	87		
baboon	146		
background correction	139	251	
back-projection principle	340		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
back-scattered electrons	367			
Bacon, Roger	53			
bandwidth	379			
barrier filter	33	67	239	255
barrier height	318			
beam expander	239	310		
beam hardening	342			
beamsplitter	27			
Beer (–Lambert) Law	38			
Bertrand lens	73			
Bessel function	37			
bias				
measurement	158	297		
process-induced	346			
bi-axially drawn film	268			
bi-modal distribution	154	202		
binary digit (bit)	88			
binding energy	366			
Binnig, Gert	360			
Biorad MRC 500	236			
birefringence	23	72		
bitmap (bmp) files				
info header for	133			
structure of	132			
blackbody radiation	40			
blending algorithm	259			
blooming	26			
Bohr, Niels	11			
quantum theory of	40			
bone	368			
Boole, Georges	89			

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links	
Boolean algebra	89		
Bragg, Sir William Henry	305		
Bragg equation	305		
‘bremsstrahlung’ radiation	332		
Brewster angle	27		
brightness	56	127	237
bulk modulus	376		
bus			
address	98		
data	98		
byte	88		

C

C80 signal processor	123		
calibration	183	249	
Z-	253		
camera (digital video)	170		
candle power	45		
capacitative array sensor	404		
carbon fibres	177		
continuous, unidirectional	375	394	
elliptical fitting to	177		
carbon tetrachloride	309		
catoptric objective	66		
Cavalieri technique	269		
Central Processor Unit (CPU)	98	131	
central stop	335		
characterisation process	130		
characterising fibre waviness	291		
charge	94	307	

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links				
charge-coupled device (CCD)	110	309	328	341	361
	369				
back-illuminated	113	336			
dark current leakage in	114				
deep-depleted	113				
defects in	116				
dynamic range of	116				
front-illuminated	112				
gain of	116				
open-electrode	113				
spatial resolution of	114				
spectral response of	112	114			
Christoffel's equations	390				
chromatic aberration	59	231			
clustering of fibres	193	222	298		
coded numbers					
binary	88				
denary	88				
hexadecimal	88				
coefficients					
geometrical	191				
reflection	24				
transmission	24				
coherence length	330				
coherent anti-Stokes Raman spectroscopy (CARS)	312				
coherent source	14	30			
collage	187				
collimated beam	29				
colour depth	127				
coma	231				
complementary probes	360				

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links				
composites	164				
carbon fibre	177	375	394		
continuous fibre	280				
fibre-reinforced	199	350			
multiple-ply	280	287			
short, glass fibre reinforced	200	260	345	366	396
unidirectional	280	345			
compound microscope	53				
compression techniques	135				
lossless	135				
lossy	135				
compressive failure	280	290			
compressive strength	279	290			
computing engine	92				
computing power	110				
concave lens	49	59			
condenser lens	61	65			
confocal image	230	328			
confocal laser scanning microscopy(CLSM)	228				
Biorad MRC 500	236				
future of	298				
NORAN Odyssey	238				
optical sectioning capability of	240				
original patents for	229				
confocal operation	310				
conjugate image	328				
connectivity	155	203	347		
4 pixel/8 pixel	206				
voxel	273	347			
contact mode SFM	316				
contrast	202	237			

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
control points	221		
convergence algorithm	287		
convergent beam electron diffraction (CBED)	370		
converging lens	50		
convex lens	49	51	59
convolution filters	77	108	144
Cornu spiral	333		
correction collars	64		
correction factors			
model-based	178		
unbiased	179		
corrections			
geometric	141	252	
light attenuation	271		
shading	249		
corrective processing	138		
correlated waviness	220		
correlation plots	391	398	
COSMOS facility	185		
counting frame			
unbiased	159	270	276
coupling fluid	56	379	
cover slip	56	61	166
critical illumination	61		
cross-bar switching	123		
crystals	23		
curvature	292		
radius of	292		
signature of	206	216	
spurious values of	291		
curvature of field	60	231	

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
D				
dark current in CCDs	114			
dark ground microscopy	69			
data bus	98			
Data Translation framegrabber card	181			
datalogging	99			
De Broglie wavelength	41	358	363	
deconvolution	249			
defects in CCDs	116			
deformation	279	324	381	
molecular	307			
Delauney triangulation	258			
denary number system	88			
density	12	376	390	
depth map	263			
depth of field	56	58	360	364
depth of focus	56			
oil immersion	243			
depth of penetration	289	298	366	
ultrasonic	376	379		
Descartes, René	86			
diascopic fluorescence microscope	67			
dichromate gelatine (DCG)	80			
dielectric edge filter	33	310		
differential interference microscope (DIC)	72			
diffraction	35			
convergent beam electron	370			
Fraunhofer	36	333		
Fresnel	333			
diffraction aperture	53			
diffraction grating	53	310	324	

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links	
diffraction pattern		
point source	52	
digital camera	170	
digital image	125	
monochrome	126	
NTSC standard	127	
structure of	126	
digital micro-mirror device (DMD)	328	
digital signal processor (DSP)	105	
digital storage scope	384	
digital volumetric imaging	223	
digitisation (of analogue signal)	116	
digitiser pad	180	202
dilation	217	
diopter	49	
Dirac, Paul	11	83
disector	217	269
look-up section of	269	
optical	270	
particulate counting in	269	
reference section of	269	
dispersion	19	115
distinct vision		
nearest distance of	51	
distortion		
geometric	141	252
diverging lens	50	
doping	351	
doubly-refracting crystal	23	76
drift velocity	83	
droplet target system	335	

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
dual path, programmable array microscope (PAM)	327
dynamic random access memory (DRAM)	95
dynamic range	
ADC	99 116 119
CCD	116 126
OCT	330
dynodes	44

E

eccentricity of lead screw	185
Eddington, Sir Arthur	11
eddy currents	404
edge effects	158
Edison incandescent light bulb	45
Edison phonograph	315
EELS technique	367
EFTEM technique	370
elastic deformation	279
elastic scattering	42
electric field	13 307 359 361 368
electric light bulb	29 45
electromagnetic radiation	13
sources	45
spectrum	14
electron back-scattered diffraction (EBSD)	365
electron beam-induced currents (EBIC)	365 371
electron energy loss spectroscopy	367
electron energy states	43
electron holography	369

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
electron microscope	332	363	
environmental scanning	365		
scanning	361		
transmission	361		
electron spectroscopy for surface analysis (ESCA)	365		
electron volt	362		
electrons	360		
back-scattered	361		
secondary emission	365		
electro-optic materials	76	312	
electro-striction	83		
elliptical fibre images	171		
fit factor for	188	204	
emission spectrum	41		
emissivity	323		
endoscope	17		
energy sensitivity	370		
Enigma codes	92		
environmental scanning electron microscope (ESEM)	361	365	
episcopic fluorescence microscope	67		
epoxy resin	166		
equilateral triangles	195		
errors			
curvature	291		
height bias	174	205	
mapping	174		
measurement	171	211	
minimising	175		
orientation	173	211	298
quantisation	119		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
errors *(Continued)*			
random	156	400	403
scanning	317		
solid angle	174	177	
stepper motor	185		
stiffness constant	394	400	
systematic	156	387	395
X-ray reconstruction	341	352	
etchants	168		
evanescent waves	383		
excitation filter	33	67	239
expectation value	120		
exposure time	257		
extended XZ sections	288		
extra-ordinary ray	23		
extruder	268		
eyepiece lens	50	54	64
field number of	64		
photo-	64		
wide-field	57		
F			
failure stress	279		
false colour	143		
fan-in/fan-out	82		
feedback	316		
shear force	320		
felt	345		
Fermat's Principle	24		
fibres			
alignment of	210		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
fibres *(Continued)*				
carbon	177	394		
characterization of	187	291	350	
clustering of	193	222	298	
crossing	217			
cross-section of	177	188		
hollow fibres	194			
intrinsic ellipticity of	203			
irregular cross-section	194			
iterative search for	288			
misalignments of	193	289		
optical	16			
orientation states of	191	208		
simulation of helical	294			
splitting of touching	206			
tracer	200	223		
triangular cross-section	194			
unidirectional	282	394		
waviness of	193	220	278	291
fibre lengths				
distribution of	219			
measurement of	212			
fibre-matrix interface	282			
fibre-optic splitter	329			
fibre orientations				
injection-moulded composite	197	351	399	
fibre strain	314			
fibre tracing	284			
automated	347			
field number	64			
field of view	57	64	157	

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
files				
bitmap-based	132			
object-based	132			
filter				
bandpass	35			
barrier	33	67	239	255
convolution	77	108	144	
correlation	108			
dielectric edge	33	310		
electronic bandpass	330			
excitation	33	67		
foam	255			
frequency domain	145			
high pass	146	237		
holographic notch	310			
interference	33			
Laplace	137	144	151	343
low-pass electronic	106	147		
low-pass optical	79			
neighbourhood averaging	144	276		
neutral density	35	68	237	
optical	33			
penetration through	256			
spatial	144	275	310	
spectral response	34			
Wien	368			
filter wheel	35			
finite element models	9	165	352	404
fish-plate	55			
fit factor	188	204	211	
flat-bed scanner	179	344		
flat-field objective lens	57	61		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
flat-fielding	138 172
flip-flops	95
flood-filling	155 194 274
flow chart	274
fluorescein	255
fluorescence lifetime imaging (FLIM)	325
fluorescence microscopy	33 67
fluorescence mode of operation	240 247 283
fluorescence process	44
inhomogeneous	271
resonance	312
fluorescent beads	233
fluorescent dye	255 326
focal length	49
zone plate	334
focal plane	
ultrasonic	378
focal plane array	325
focus	28
force-distance curve	315
forensic science	313
Fourier Transform	78 83 145 325 374 380 389
Fast (FFT)	145
Fourier Transform Raman spectroscopy	313
fractal coefficients	192
fracture	279
frame averaging	139
framegrabber	97 123 181
Data Translation	181
Fraunhofer diffraction	36 333

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links				
frequency	11				
nuclear resonance	371				
frequency distribution					
number-weighted	278				
volume-weighted	278				
Fresnel diffraction	333				
Fresnel zone plates	333				
frog-plate	55				
FT-IR technique	322				
future scenarios	405				
fuzzy logic	92				
fwhm (full width half maximum)	236				
G					
gain					
CCD	116				
PMT	273				
variable	274				
Galileo, Galilei	55				
gallium nitride (GaN)	368				
galvanometer	236	240	253		
Gauss plane	362				
gelatin	33				
Genesis framegrabber	123				
geometric correction	141	252			
geometric distortion	141	252			
geometrical optics	51				
glass fibre reinforced composites	200	260	345	366	396
glycerin	68				
GOI (Kentech Instruments)	327				
GPIB serial link	384				

This page has been reformatted by Knovel to provide easier navigation.

Index terms **Links**

graticule 252
gravity wave detection 32
group velocity 17
gyromagnetic ratio
  hydrogen 373

H

Harvey, William 55
height bias 174
Heisenberg, Werner von 11 156
Herman's orientation parameter 191
Herschel, William 324
hexadecimal codes 88
high angular annular dark field imaging
    (HAADF) 369
high-order polynomials 291
high-pass filter 146
histogram
  bi-modal intensity 154 202
  equalisation 143
  inversion 143
  logarithmic 143
  manipulation of 142
histostack 271
holographic materials 80
holographic notch filter 80
Hooke, Robert 55
Hooke's Law 381
Hough, Paul 147
Hough Transform 147 214 343
  circular 152 202

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links
Hough Transform *(Continued)*		
generalised	152	
linear	147	215
parametric form of	148	215
HREM	364	
human eye	48	
human eye response	38	
Huygens, Christian	10	35
Huygens secondary wavelets	35	333
hydrogen emission lines	41	
hypercube network	104	
I		
illumination		
critical	61	
epi-	229	
Koehler	62	67
Nipkow disk	231	
optical fibre	17	
problems with	172	
uniform	45	63
image		
brightness of	56	
calibration of	183	
characterization of	130	156
classification of	130	156
confocal	231	
contrast of	166	202
digital	125	
distortion of	141	
elliptical	171	
enhancement of	137	

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links	
image *(Continued)*			
format of	134		
frame averaging of	139		
geometric correction of	141		
height of	262		
intermediate	57	360	
real	50		
reference	249		
segmentation of	152		
shadow	339		
storage and retrieval of	131		
structure of	132		
topological	316		
virtual	50		
image acquisition	129		
image analyzer design			
final Leeds (LAAM)	179		
initial Leeds transputer	182		
image artefacts	274	341	
image aspect ratio	173		
image characterization	129		
image contrast	202		
image compression	135		
image distortion	8	141	252
measuring	252		
image enhancement	129	137	
image manipulation	130		
image noise	274	285	
image plane	56		
image processing	8	129	
implementation of	131		
image segmentation	129	152	

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
image voxel	271
imaging modes	254
auto-focus	265
fluorescence	254 283
fluorescence lifetime	325
high angular annular dark field	369
NMR	375
photon counting	257
reflection	254 288
immersion medium	
optical	243 247
ultrasonic	384
included angles	191 193
incoherent light source	14 30
indium antimonide	323
inelastic collisions	361
infinity corrected objective	60 170
information channel	363
‘Information Theory’	92
inhomogeneous fluorescence	271
injection-moulded composites	197
orientations of	201
push-pull	209 396
INMOS	
A100 dsp	106
T800 transputer	103 182
instruction sets	131
instrumental noise	293 296
intensity	
image	233
light	24

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
intensity gradient function	286			
interfaces	377			
propagation across	401			
interference	29			
constructive	32			
destructive	32			
division of amplitude	30			
division of wavefront	30			
interference fringes	282			
interferometer				
Michelson	32	185	324	329
interferometry	32			
lamellar grating	325			
wavefront modulation	70			
wavefront shearing	70			
Internet control of microscopes	368			
intrinsic fibre ellipticity	203			
iodine doping	351			
ionisation	44			
ionisation energy	366			
IR vibro-thermography	324			
iris diaphragm	65			
irradiance	20	29		
irregular cross-section fibres	194			
isotropy	210	382		
test for	179	298		
iterative fibre search algorithm	286			
iterative process	393			

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
J	
JPEG standard	137
artefacts in	138
judgement	292
K	
kernel	107 144
Kerr cell	76
Kerr effect	76
kink bands	279
Kirkpatrick-Baez mirror	333
Koehler illumination	62 67
L	
Labview (National Instruments)	100 405
lamellar grating interferometry	325
Laplace filter	137 144 151 343
large angle sectioning	175 281
large area automated microscope (LAAM)	179 181 190
laser	45 47 238
tuneable	312
lateral magnification	50
lateral resolution	233 245
latex spheres	146 265
lattice dynamics	313
lead screw eccentricity	185
least distance of distinct vision	49 51
least square fitting process	187 189
Leitz, Ernst	60
lens	
achromatic doublet	58

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links	
lens *(Continued)*			
Bertrand	73		
concave	49		
condenser	61		
converging	50		
convex	49		
diverging	50		
electrostatic	359		
eyepiece	50	54	64
magnetic	359		
special tube	60		
transformation properties of	77		
lens maker's formula	49		
light collecting power	56		
light dependent resistor	91		
light emitting diode (LED)	83		
light intensity	24		
light rays	24		
light velocity	13		
line detection	151	213	
line profiles	251	388	
Lissajous figures	21		
Lister, Joseph Jackson	59		
logical operations	89		
AND, OR	91		
longitudinal pressure waves	390		
look-up tables	187		
Lorentz microscopy	369		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
M	
machine code	102
macromolecules	267
magnetic field	13 359 361 368 371 373
magnetic moment of proton	371
magneto-optic materials	76
magnification	
empty	58 65
lateral	50
useful	58
variation of	253
zone plate	336
magnifying lens	48
magnifying power	
compound microscope	55
simple microscope	50
telescope	54
Malpighi, Marcellus	55
Malus' Law	23 73
mapping	385
marching cubes	258
marginal rays	244
matrix in composites	154
acid digestion of	213
autofluorescence of	246
pyrolysis of	213
Matrox framegrabber	125
maximum useable depth	246
Maxwell, James Clerk	11 12 39
em equations	13 27

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
measurement bias	158	211	297	
curvature	291			
mechanical switch	90	92		
Melinar disk	245			
membership criterion	155			
memory				
dynamic random access	94			
read only	95	96		
static random access	94	95		
video-RAM	94			
mesh array	104			
mesostructures	6	165	391	
metallisation of surfaces	264			
Michelson interferometer	32	185	324	329
microprocessor	98			
microscope				
atomic force	364			
compound	53			
confocal laser scanning	231			
diascopic fluorescence	67			
dual path, programmable array (PAM)	327			
episcopic fluorescence	68			
high resolution, electron	363			
metallurgical	56			
nano-engineered, optical	406			
Nomarski interference	71			
Olympus BH2	164	168		
optical reflection	163			
polarizing	73			
reflecting optics	66			
scanning electron (SEM)	364			
simple	48			

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links
microscope *(Continued)*		
smallest	83	
stage-scanning	231	
transmission electron (TEM)	332	361
X-ray	331	
microscopy		
atomic force (AFM)	314	319
brightfield	65	
confocal laser scanning	231	
darkfield	65	
darkground	69	
electron	361	
fluorescence	33	67
interference	69	
Lorentz	369	
multi-modal	306	
near-field	314	
near-field scanning optical	319	
optical reflection	66	163
optical transmission	65	
phase contrast	69	
polarization	73	
Raman	307	
scanning acoustic (SAM)	378	
scanning force	315	
scanning probe	314	
scanning thermal	317	
scanning tunneling (STM)	317	
two-photon excitation	298	
UV Raman (UV RS SNOM)	322	
X-ray	331	
microtoming	179	230

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
microtomography			
3D ion	404		
X-ray	331		
‘Milliscope’	320		
minimisation routine	393		
Minski, Marvin	228		
MIPs	110		
mirror surface	264		
mirrors	49		
galvanometer driven	236	240	253
Kirkpatrick-Baez	333		
micro-miniature	328		
optically flat	24	235	250
misalignment angle	279	289	
ply	193	291	
missing data	349		
modulation envelope	18		
modulators	76		
moments technique	187		
equations for	203		
monochromatic beam	366		
monochromatic light	35	307	
Monte Carlo methods	368		
Morse Code	87		
Motorola 6802 integrated circuit	97		
multiplex advantage	325		
multi-ply composite	287		
multiply-add and delay cell	106		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
N				
nano-engineered microscope	84	406		
National Instruments Inc.	405			
near-field microscopy	314			
near point of the eye	49	51		
nearest neighbours	252	298		
needle-punching	345			
neighbourhood averaging	264	276		
neighbourhood operations accelerator	124			
neutral density filter	35	68	237	310
neural networks	80			
adaptive	81			
dynamic holographic	80			
optoelectronic	80			
neutron radiography	404			
neutrons	360			
Newton, Sir Isaac	10	28		
nibble	88			
Nikon Optiphot	67			
Nipkow disk	231			
noise				
dark	115			
image	172	274	285	
instrumental	293	296		
mechanical	389			
Poisson	364			
random	139			
shot	364			
spatial	274			
noise function (simulated)	293			
Nomarski interference microscope (NIC)	71			

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
non-linear signal processing	76			
NORAN 'Odyssey' CLSM	238			
NSOM/SNOM	320			
NTSC standard video	127			
nuclear magnetic resonance (NMR)	371			
applications of	375			
number system				
binary	88			
denary	88			
hexadecimal	88			
numerical aperture(NA)	53	56	233	241
inscribed	60			
oil immersion objectives	246	262	330	
Nyquist sampling theorem	122			
O				
object-based storage	132			
objectives				
acoustic	378			
air	63	267		
built in iris diaphragm for	65			
catoptric	66			
correction collars for	64			
depth of focus of	243			
DIC	72			
electron	360			
flat-field	57	61		
infinity-corrected	60	170		
metallurgical microscope	56			
NIC	72			
oil immersion	56	63	262	266
Olympus	170			

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links	
objectives *(Continued)*			
parcentric	64		
parfocal	64		
planachromats	61		
planapochromat	61		
solid angle of	264		
ultrasonic	378		
working distance of	63	246	
OCCAM language	103	182	
ocular	50	54	
oil immersion objectives	56	243	
Olympus BH2 microscope	163	168	
operating system (OS)	132		
optical axis	37		
optical cantilever	306		
optical computer	73		
optical coherence tomography	298	329	
optical density (OD)	38		
optical flat	24		
optical fibres	16		
single-mode	320		
tapered	320		
optical filters	33		
optical pathlength	15		
optical reflection microscopy	66	163	
optical sectioning	230	240	
optical sectioning method (OSM)	223		
optical transmission microscopy	65		
optical transparency	38	248	
optical units	233		
optoelectronic neurocomputing	80		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links				
orientation distributions	174	176	190	198	207
orientation states	191				
oxidation process	167	202			

P

packing fraction					
fibre	207	247	283	348	
pan and zoom	249				
parallel processing	105	123			
parameter space	150				
parametric equation	148				
paraxial rays	28	244			
parfocalising distance	63				
particle accelerators	331	333			
particle momentum	358				
particulate analysis	256				
number-weighted	277				
thin film	267				
unbiased	270				
volume-weighted	277				
pattern matching					
2D sections	218	223	262		
equations for	221				
Peltier cooling system	112	114	310	323	
perimeter tracing	207				
Peripheral Component Interconnect (PCI)	124				
peripheral interface adaptor (PIA)	99				
permeability	14				
permittivity	14				
Perspex (PMMA)	389				
stiffness constants of	394				

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
petri dish	218
phase	12
phase contrast microscopy	69
phase difference	15 333
phase encoding gradient	374
phase plate	70
phase transition temperature	317
phosphorescence	44
photobleaching	68 238 256 298
photocathode	44
photoelectric effect	39 44 110
photo-eyepiece	64
photo-ionisation process	365
photomicrograph	60 201
photomultiplier tube (PMT)	44 236 240 264 271 361
photon	39 371
photon counting mode	45 257
photonics	73
photo-refractive materials	80
physical sectioning	230
pictograms	87
picture elements (pixels)	111 121 126 173
piezoelectric scanner	238 253 316 318
piezoelectric transducer	377
pinhole	52 242
pipelines	104 106
planapochromat objectives	61
Planck, Max	39
Planck's constant	39 41 358 371
ply boundaries	289

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
ply misalignments	193 280
Pockels Effect	76
point detector	45 325
point scanning	236
point source	52
point spread function (PSF)	
confocal microscope	233 244 245
conventional microscope	233
Poisson noise	364
polarisability	307
polarization	307
state of	20 377
polarization microscopy	73
polarized light	
circularly	22
elliptically	21
linear (plane)	20
un-	20
polarisers	22 72
Polaroid film	22
polishing samples	166
pollen grains	230
polyamide66	337
polyarylamide non-woven felt	345
polycrystalline materials	333
polyethylene (LDPE) film	271 275
polymer films	267
characterization of	268
surface profile of	266

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links	
polymers			
conjugated	310		
refractive index of	284		
polymethylmethacrylate (Perspex)	389		
polynomial function	292	349	
polypropylene	337		
glass fibre-reinforced	398		
polyurethane foams	255	260	267
position vector	292		
positron annihilation spectroscopy	404		
potential energy	371		
potting samples	166		
power (of lens)	49		
power spectral density	291		
Poynting vector	20		
precision (of surface height)	262		
precision advantage	325		
principal rays	50		
priority encoder	119		
Prior Scientific XY stage	170	181	
prism (electrostatic bi-)	369		
probe tip	315	318	
oscillating	320		
profilometers	315		
propagation			
light wave	24		
ultrasonic	383		
X-ray propagation	331		
proton magnetic moment	371		
push-pull processing	209		
pyrolysis	213		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
Q				
quanta	39			
quantisation error				
ADC	119			
quantum computer	109			
quantum interference	109			
quantum mechanics	40	309	317	372
quantum optics	83			
quantum theory	39	358		
quarter wave film	379			
quarter wave plate	22			
quartz	378			
quasi-longitudinal wave	377			
quasi-transverse wave	377			
R				
radius of curvature	49	292		
frequency distribution of	296			
Raman, C.V.	43	307		
Raman effect	307			
resonance	310			
stimulated	312			
Raman microscopy	307			
Raman Scattering	43	308		
Raman shifts	311	313		
Raman spectroscopy				
surface enhanced	313			
Ramanscope	309	321		
random access memory				
dynamic	95			
static	95	123		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
random access memory *(Continued)*	
video	95 97
random errors	400
raster scanning	186 257
Rayleigh, Lord	5 52
Rayleigh criterion	52 233 262 305 358 363 364
Rayleigh scattering	42 238 308 380
rays	24
marginal	244
paraxial	28 244
read only memory (ROM)	95
electrically eraseable (EEROM)	95
eraseable, programmable (EPROM)	100
real image	50
reality	
physical	86
virtual	86
recombination algorithm	273
reconstruction	
volume	257 272 296
X-ray	340
reduced instruction set chips (RISC)	103 123
reference image	249
reflection	
coefficient of	24 25 39 167
diffuse	25
specular	24
refracting telescope	53
refraction	14 390

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
refractive index	14 49 76 80 243 247 263 268 284 331 338
region growing	155
remote control	405
renderings	
surface	258
volume	258
repeatability of XY positioning	185
reproducibility	
composite structure	400
time-of-flight	388
resolution	
axial	234 245 250 262 375 379
lateral	233 245 325 359 361 379
spectral	309
ultrasonic	394
X-ray	341
resonance absorption	42 371
retina of the eye	48
RGB colour mode	127
Rhodamine dye	255 275
Roentgen, Wilhelm	337
Rohrer, H.	360
RS170 video waveform	121
RS232 interface	99
RS432 interface	384
run length encoding (RLE)	135
Ruska, Ernst	360

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
S				
sample preparation	166			
cutting	166			
oxidation	167			
polishing	166			
potting	166			
sapphire	378			
satellite objects	274			
scalar product	371			
scanning				
piezoelectric	238	253	316	319
point	236			
stage	231			
videorate	238			
scanning acoustic microscopy (SAM)	376	378		
scanning galvanometer	236			
scanning probe microscopy	314	360		
scanning thermal microscopy	315			
scanning tunneling microscopy (STM)	315			
scattering cross-section	39			
scattering	39			
elastic	42			
inelastic	43	361		
scintillator	341	361		
second order intensity function	193			
secondary emission electrons	365			
secondary ion mass spectrometry (SIMS)	404			
secondary wavelets	35	333		
sectioning				
optical	261			
physical	269	281		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links				
sectioning schemas	261				
XY at different Z	261	282			
XZ at different Y	262	284	286		
seed point	155				
segmentation	152	274	285		
semiconductors	313	321	364	369	370
sensor	91	99			
shading	139				
shading correction	249				
shadow images	337	339			
shear force feedback	320				
shear modulus	12	376			
shear pressure waves	12	383			
velocity of	12	390			
sign convention (lens)	50				
signal enhancing lens system (SELS)	298				
signal to noise ratio (S/N)	139	374			
silicon carbide mats	166				
simulation of helical fibres	294				
single-pass algorithm	187				
skeletonisation	343	347			
skewing of images	141				
skin-core-skin structure	210	218	350	398	
slow-scan ADC	119				
smallest confocal microscope	83				
Snell's Law	15	387			
SNOM/NSOM	320				
Sokilov, S.Y.	378				
solid angle effect	174	177			
sources	45				
coherent light	14	30	46		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
sources *(Continued)*				
confocal light	231			
incoherent light	14	30		
space curve method	291	347	350	
curvature from	292	350		
torsion from	292	350		
spatial filter	70	78	144	310
problem with	275			
spatial frequency	362			
spatial light modulator (SLM)	82	327		
spatial noise	274			
spatial resolution	52			
CCD	114			
circular aperture	37			
conventional microscope	7	24	376	
electron microscope	359	361	363	
NMR	375			
ultrasonic	376	379		
spectral line				
width of	375			
spectral resolution	309			
spectral response				
CCD	112	114		
filters	34			
fluorescein dye	255			
spectroscopy				
Auger electron (AES)	365			
coherent anti-Stokes Raman	312			
electron energy loss (EELS)	367			
positron annihilation	404			
Raman	307			
X-ray photoelectron	365			

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
specular reflection	24			
angular dependence of	27			
spherical aberration	39	243	244	247
spike pulser	384			
spin echo method	374			
splitting images	206			
spontaneous emission	43			
spurious curvature	291			
spurious peak	174			
sputter coating	264	276		
STAMP microcontroller	384			
Stefan-Boltzmann Law	322			
stepper motors	35	384		
principle of	183			
repeatability of	185	253	262	
stereology	8	156	177	213
correction for height bias	205			
particulate	269			
stiffness constants	380	382	390	
equations for 3D	390			
maps of	402			
measurement scheme for	393			
Perspex	394			
stimulated emission	45			
Stokes Raman signal	308			
Stokes shift	39			
stopping potential	40			
strain				
coefficients of	381			
definition of	381			
strain matrix	381			

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links				
stress					
definition of	381				
failure	279				
matrix shear yield	279				
real-time	352				
stress profile	212				
stress tensor	381				
stretching images	141				
String Theory	9				
structure					
atomic	40				
meso-	6	10	165		
micro-	6	167			
nano-	6	10	84	406	
skin-core-skin	210	218	350	398	
superconductor	370				
surface acoustic waves	377	383			
surface gradient	264				
surface rendering	258				
surface studies	261	264	266	316	364
surface topology	71	264	266	316	364
switches					
electronic	89				
mechanical	90				
optoelectronic	76				
time-gated	327				
symmetry axes	382	393			
synchrotron radiation	333				
systematic errors	156	387	395		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
T			
tapping mode AFM	319		
TE mode	27		
telescope			
refracting	53		
tensile pressure waves	377		
velocity of	385		
tensor coefficients	192		
fibre orientation	191	208	350
textiles	343		
non-woven	350		
woven	345	350	
thermo-electric cooler	114		
thermography	322		
thin films			
biaxially-drawn	267		
particulate analysis of	267		
polymer	267		
surface structure of	66	364	
thresholding			
circuit for	91		
image	154	202	
throughput	80		
throughput advantage	325		
tilt angle	264	328	
time constants			
nuclear relaxation	374		
RC	107		
time of flight system	379	384	
TM mode	27		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links		
tomography			
optical coherence	298	329	
X-ray micro-	331		
torsion	292		
total internal reflection	16		
tracer fibres	200	223	
transducers			
immersion	384		
piezoelectric	253	378	384
transfer function			
optical system	79		
transformation equations	176		
transformation function	142	252	
transistor			
age of the	92		
field effect	93		
invention of the	95		
transmission axis	23		
fast	23		
slow	23		
transmission coefficient	24	39	
transmission electron microscope (TEM)	361		
transparency			
optical	38	248	
transputers (T800)	103		
network of	104	182	
transverse pressure waves			
velocity of	12		
transversely isotropic material	382	394	
stiffness constants of	382		
triangular cross-section fibres	194		

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
truth tables	91
tube length	56
tube lens	60
tunnelling current	317
Turing, Alan	87
twist	345 351
two-photon excitation microscopy	298
U	
unbiased brick	270
unbiased counting frame	159 270 276
Unicode	87
unidirectional fibres	282
ultrasonic testrig	383
response of	394
stiffness constant equations for	390
ultrasonics	329 376
Ultraviolet Catastrophe	40
unit vector	208
V	
van Deijl, Harmanus	58
van Leeuwenhoek, Antoni	55
variable gain technique	271 276
vector matrix multiplier	82
vectors	13 20 258 347 371 383
position	292
VELA	99 101 405

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
velocity	
harmonic	11
light wave	13 15
longitudinal wave	386
shear wave	13 386
temperature dependence of	388
ultrasonic	386 390
velocity selector	368
veto operation	248
vibrations	318
video camera	
Sony XC-77 digital	180
video-conferencing	405
video interface chip (VIA)	124
video RAM memory	97 181
video signal	121
Vikings	49
virtual image	49
virtual instrument	405
virtual screen	258
Visser correction	243
visualisation (in 3D)	10 257
voids	154 375 380
volume element	382
volume reconstruction	257 288 296
volume rendering	258
von Neumann machine	87
Voronoi cells	193
voxels	
attenuation of	271
connectivity of	273 347

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links			
voxels *(Continued)*				
intensity of	259	348		
NMR	374			
W				
wave equation	11	383		
wave vector	20			
wave velocity	11	385		
wavefront modulation interferometer	70			
wavefront shearing interferometer	70			
wavelength	11			
De Broglie	41	363		
fibre	291			
X-ray	291			
wavenumber	12	17		
waves				
electromagnetic	11			
longitudinal (tensile)	390			
transverse (shear)	390			
waviness				
case study in	278			
characterizing	291			
correlated fibre	220			
weighting function	178			
wide-field eyepiece lens	57			
Wien filter	368			
Wien's Law	40			
Wollaston prism	71			
work function	44			
working distance	63	67	246	331

This page has been reformatted by Knovel to provide easier navigation.

Index terms **Links**

X

X-ray crystallography 305
X-ray emissions
  characteristic lines in 366
  energy dispersive (EDX) 365
  wavelength dispersive (WDX) 365
X-ray micro-radiography 336
X-ray microscopy 331
  soft 333
X-ray microtomography 337
  case study for 337
  dopant for 351
X-ray mirror
  Kirkpatrick-Baez 333
X-ray photoelectron spectroscopy (XPS) 365
X-ray tube 332 337 341
X-rays 360
  attenuation of 331
XY sections 261 282
XY stage (Prior Scientific) 181
XZ sections 262 282
  extended 288

Y

yarn 345
Young's Modulus 376 380 381
Young's Slit experiment 32

Z

Z-calibration 253
Zadeh, Lofti 92

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
Zeeman effect	371
Zeiss, Carl	60
Zernike, F.	70
zinc oxide (ZnO) transducer	378
zone plates	333 339

This page has been reformatted by Knovel to provide easier navigation.